Children's learning of arithmetic facts

van Galen, M.S.

2009

Children’s learning of arithmetic facts

ISBN/EAN: 978-90-8659-341-5
Printed by Ipskamp, Enschede
Distributed by PI Research, Postbus 366, 1115 ZH Duivendrecht

VRIJE UNIVERSITEIT

Children’s learning of arithmetic facts

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op dinsdag 22 september 2009 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Mirte Susanne van Galen

geboren te Arnhem

promotor: prof.dr. P. Reitsma

Contents

1

General introduction

A common observation among primary school teachers is that some children have difficulties in learning simple arithmetic facts like 2 + 3 = 5 or 3 × 4 = 12. Although most children seem to develop representations of simple arithmetic facts in long-term memory almost automatically, there are other children who need to count to find the answer to 2 + 3 even after years of mathematics education. This thesis focuses on the process of automatization in simple arithmetic. In the present research, a broad approach towards automatization is taken. Some may say that automatization is only accomplished when a child uses retrieval from long-term memory, but this definition is probably too rigid. Recent research has established that even adults regularly use nonretrieval strategies for simple arithmetic problems (Campbell & Austin, 2002; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). In this thesis, automatization is defined as gaining efficiency in solving simple arithmetic problems, which means that problems are solved faster and less errors are made. There may be several reasons for more efficient problem solving, such as faster procedural skills, a transition from counting to retrieval, or faster retrieval. In the present research, we aim to answer several questions that are related to automatization in simple arithmetic. What is a good practice method for children to gain efficiency in solving simple arithmetic problems? How are arithmetic facts represented in long-term memory? Is there a way to predict which children are good and which children are poor in learning arithmetic facts, for instance from their ability to retain information in short-term memory? Is the ability to solve simple arithmetic problems related to numerical estimation? Furthermore, the automatization of basic numerical processes is investigated: how do children develop automatic access to number magnitude? These questions will be treated in the next chapters. First, in this general introduction an overview of arithmetic fact learning in children and in adults is presented.

1.1 Typical development in simple arithmetic

Initially, children solve simple addition problems with counting procedures (either finger counting or verbal counting). The most simple counting procedure is termed ‘counting all’ and involves counting both addends, starting from 1 (e.g., 2 + 3 = 1, 2, 3, 4, 5). A more sophisticated strategy is counting on from the first addend (2 + 3 = 2, 3, 4, 5). Even more efficient is counting on from the largest addend (2 + 3 = 3, 4, 5). However, in contrast to counting procedures, usually the fastest strategy is to retrieve the answer from long-term memory (2 + 3 = 5). Apart from counting and retrieval strategies, children also use decomposition strategies, which involves breaking up the problem into more simple problems; for example, 8 + 7 can be decomposed into [8 + 2] + 5. At each point of mathematical development, children have several strategies available. The strategy that is chosen for a specific problem depends on the experience of the child and the characteristics of the problem. In other words, a child may still use 'counting on from first' when it also knows the strategy of 'counting on from larger' and it can use retrieval on some problems and counting on other. There are no sudden transitions from one strategy to another, but instead gradual shifts occur towards a more mature mix of strategies (Siegler, 1987, 1996). Through extensive practice with addition problems, children in primary school move on from initial counting procedures to faster memory-based strategies (Geary, Bow-Thomas, Liu, & Siegler, 1996; Goldman, Mertz, & Pellegrino, 1989; Siegler, 1987). A comprehensive study with single-digit addition problems reports the following percentages on the use of retrieval in American children: 20 % of the trials at the beginning and 28 % at the end of Grade 1; 31 % at the beginning and 41 % at the end of Grade 2; 45 % at the beginning and 56 % at the end of Grade 3 (Geary et al. 1996).

The development in multiplication is similar to addition. Initially, children use for instance repeated addition (e.g., 3 × 4 = 4 + 4 + 4) or counting sequences (e.g., 3 × 5 = 5, 10, 15) to solve simple multiplication problems, but eventually retrieval is the most frequently used strategy (Cooney, Swanson, & Ladd, 1988; Mabbott & Bisanz, 2003; Lemaire & Siegler, 1995). A study in the United States with single-digit multiplication problems reports the use of retrieval in 55 % of the trials in Grade 3 and 74 % in Grade 4 (Cooney et al., 1988). A similar study in Canada reports 67 % retrieval in Grade 4 and 88 % in Grade 6 (Mabbott & Bisanz, 2003). Of course, an increased reliance on fact retrieval is not the only source of improvement in simple addition and multiplication. Accuracy and solution times can also be improved by better procedural skills or faster retrieval.

1.2 Children with mathematical difficulties

One of the most consistent findings in the literature on children with mathematical difficulties is that they often have particular problems in memorizing basic arithmetic facts. Having said this, the issue is first how we define mathematical difficulties. There are several terms one could use to refer to mathematical problems in children: for instance, mathematical disabilities, mathematics disorder, or developmental dyscalculia. In this introduction I will use the more neutral term 'mathematical difficulties' to indicate that I refer to the broad category of children that experience severe mathematical problems, but who may not all have an actual disorder. The DSM-IV, a clinical classification scheme for psychiatrists and psychologists, uses the term ‘mathematics disorder’ and describes this learning disorder as having a mathematical ability substantially below what is expected, given the person’s age, intelligence, and education. Educational psychologists would add that the mathematical problems should also be persistent over some years before classifying a child as having mathematical difficulties. Note that the before-mentioned criteria are purely observational in nature and do not state anything about the cause of the learning problems.

Not only is there confusion in the literature about which term to use, also very different criteria are used to define mathematical difficulties. Some studies use a very strict cutoff-score to select children with mathematical difficulties, whereas other studies select a much wider group of low achieving students, for instance children with scores under the 25th percentile of a standardized test for mathematical achievement. Because of these different criteria, there is a large variation in the estimated prevalence of mathematical difficulties; estimates range from 3 % to 8 % in the school population (Geary, 2004; Shalev, Auerbach, Manor, & Gross-Tur, 2000; Shalev & Gross-Tur, 2001). Of course, one should keep in mind that mathematical abilities, like all abilities, lie on a continuum, and that a cutoff-score is always more or less arbitrary. Furthermore, children with mathematical difficulties form a heterogeneous group, i.e., they may show different kinds of mathematical problems. These problems include for example conceptual problems with the base-10 numeral system, frequent errors in the execution of procedures, and difficulties with retrieving arithmetic facts.

Even though definitions of mathematical difficulties may vary, research over the last decades has made significant progress in investigating the mathematical development of children with mathematical difficulties. Many children with mathematical difficulties have a poor conceptual understanding of counting principles; for example, they do not recognize that the order in which objects are counted is irrelevant (Geary, Hamson, & Hoard, 2000; Geary, Hoard, & Hamson, 1999). Furthermore, they use a less mature mix of strategies than their normal developing peers when solving simple arithmetic problems. For instance, they rely longer on finger counting and they use the counting-all procedure more frequently (Geary et al. 1999; Geary et al., 2000). Finally, these children often have particular difficulties in retrieving basic arithmetic facts from long-term memory (Andersson, 2008; Geary, 1993; Hanich, Jordan, Kaplan, & Dick, 2001; Russell & Ginsburg, 1984). For most children with mathematical difficulties the ability to retrieve arithmetic facts does not appear to improve substantially during elementary school (Jordan, Hanich, & Kaplan, 2003; Ostad, 1998). Research over the past years has established that children with mathematical difficulties often have working memory problems, which may in part explain their problems in learning mathematics (D’Amico & Guarnera, 2005; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; McLean & Hitch, 1999; Passolunghi & Siegel, 2004; Rosselli, Matute, Pinto, & Ardila, 2006; Siegel & Ryan, 1989).

Whether children with mathematical difficulties have fundamental problems in mathematical processing or are merely delayed in their mathematical development is still under discussion. Russell and Ginsburg (1984) compared fourth-grade children with mathematical difficulties not only to normal developing children of the same age, but also to a group of third graders. They found that children with mathematical difficulties are in many respects similar to normal developing, younger children, thus indicating that their development is delayed but not fundamentally different. However, there was one dramatic exception, namely recalling basic arithmetic facts. Although the study of Russell and Ginsburg suggests that children with mathematical difficulties are mainly delayed in their development, over the last decade evidence has been found that children and adults with mathematical difficulties may be impaired on some domains of basic numerical processing (Bachot, Gevers, Fias, & Roeyers, 2005; Landerl, Bevan, & Butterworth, 2004; Rouselle & Noël, 2007; Rubinsten & Henik, 2005). Geary (1993, 2004) distinguishes three different subtypes of difficulties in mathematics. The procedural subtype, which involves difficulties in executing mathematical procedures, appears to represent a developmental delay. In contrast, the semantic memory subtype, associated with difficulties in retrieving arithmetic facts, appears to represent a true developmental difference from normal developing peers. This subtype seems to co-occur with phonetic forms of reading disabilities and might be a heritable deficit. The third subtype, the visuospatial subtype, is associated with difficulties in spatially representing mathematical information, but much is still unclear about this subtype.

1.3 Network models of arithmetic fact retrieval

There has been a long history of studying arithmetic fact retrieval in adults, especially in multiplication. Research has revealed several robust effects in arithmetic performance. The most notable effect is the problem-size effect: problems with large digits and large answers in terms of number magnitude are more difficult to solve (response times are longer and more errors are made) than problems with small digits and small answers. Other effects are the tie effect (an advantage for problems with two identical operands, for instance $2 + 2$ or 6×6) and the five effect in multiplication (an advantage for problems with 5 as a factor, for instance 3×5). Also, an important observation in arithmetic performance is that errors are not random. Most errors in multiplication are table related; an example of such an error is $6 \times 4 = 28$ instead of 24. Furthermore, table related errors are characterized by a distance effect: responses are more likely to be correct answers for problems nearby in the multiplication tables than for problems further away.

The observation that errors are often table related has led researchers to believe that the knowledge of arithmetic facts is organized as an associative network in long-term memory. In such a network various connections exist between a problem and possible answers. When a problem is solved correctly on repeated occasions, the association between the problem and the correct answer becomes stronger, that is, the probability for correct retrieval increases. This idea of an associative network is corroborated by studies on interference and priming. Campbell (1987) found that retrieval of the product of a multiplication problem increases the probability that this product will be retrieved in error to another problem later, especially when there is a relatively strong association between the problem and the false answer that is primed. For instance, the probability of responding "56" to 7 × 9 is higher if 56 has been retrieved recently via 7 × 8. Galfano, Rusconi, and Umiltà (2003) showed that the presentation of two numbers, e.g., 6 and 4, automatically activates not only the corresponding multiplication problem (6 × 4) and its correct response (24), but also the immediate neighbours of that multiplication problem (5 × 4 = 20, 7 × 4 = 28, 6 × 3 = 18, 6 × 5 = 30). The findings of Galfano et al. provide strong evidence that multiplication facts are stored in a highly related network in which activation spreads from the product node to adjacent nodes. In a later study (Rusconi, Galfano, Rebonata, & Umiltà, 2006) it was found that the links between operand and product nodes are bidirectional; it is not only true that presenting two numbers leads to an activation of the product node, but presenting the product of a multiplication problem (e.g., 24) also leads to activation of the operand nodes (e.g., 6 and 4).

There have been various proposals for the architecture of an associative network for arithmetic facts. There is the network retrieval model of Ashcraft (1987), the connectionist model Mathnet of McCloskey and Lindemann (1992), the network interference model of Campbell (1995), and recently Verguts and Fias (2005) presented their interacting neighbours model. However, Siegler (1988) is the only researcher who attempted to integrate the mathematical development of children in his distribution-of-associations model. This model will be discussed in the next section.

1.4 Arithmetic fact retrieval in children

Compared to the large number of studies on arithmetic fact retrieval in adults, much less research has been conducted on arithmetic fact retrieval in children. Lemaire, Barrett, Fayol, and Abdi (1994) found evidence that children as young as seven years old activate addition facts automatically. When asked to indicate whether a number had been present in a previously viewed number pair, children from different ages took more time to reject distractors equal to the sum of the pair than unrelated numbers. However, the interference effect depended on the size of the numbers in the pair and the age of the child. Seven-year-olds showed an interference effect only when both integers in the original pair were 5 or smaller (small problems), whereas eight-year-olds also showed an interference effects when one integer was 5 or smaller and the other was between 6 and 9 (medium problems). Nine- and ten-year-olds showed an interference effect on all presented problems, including large problems in which both integers were between 6 and 9. Also, in the same study, a verification experiment with addition and multiplication problems showed that children were slower in responding to problems with false answers that would be correct with the other operation (e.g., 3 + 5 = 15 or 3 × 5 = 8) than to problems with other false answers. Again, this interference effect depended on the size of the integers and the age of the child. In multiplication, eight-year-old children were found to be confused by the sum of the integers (3 × 5 = 8), but only on small problems. With increasing age, children did not only show a confusion effect on small problems, but also on medium problems. In addition problems, a confusion effect (3 + 5 = 15) was not present until children were nearly nine. At this age, the confusion effect was found only on small problems, whereas in nine- and ten-year-old children it was found also on medium problems. The findings of Lemaire et al. (1994) are consistent with associative models of arithmetic fact retrieval and suggest that, throughout primary school, children gradually expand their associative network for arithmetic facts.

How do children develop an associative network for arithmetic facts? Siegler (1988) assumed that children initially solve multiplication problems by nonretrieval strategies, such as repeated addition. The answers found with nonretrieval strategies (both correct and incorrect) shape the later associative network. If sufficient associative strength has accumulated between a problem and one or more answers, the answer is retrieved directly from memory. This is not to say that direct retrieval always leads to the correct answer, because there may be interference from competing false solutions. However, with practice and feedback, "peaks" will develop in the distributions of associations such that the strongest associations are between arithmetic problems and their correct answer. In Siegler's distribution-of-associations model, when a child attempts to solve a multiplication problem, first the confidence criterion is set. The confidence criterion is a value that must be exceeded by the associative strength of a retrieved answer for the child to state that answer. There is also a second parameter, the search length, which indicates the maximum number of retrieval efforts that the child will make. If the child does not succeed in retrieving the answer from memory within the permissible search length, it will use what Siegler calls a "backup strategy", i.e., a nonretrieval procedure.

The distribution-of-associations model has been criticized for focusing solely on retrieval and on item-specific practice and thereby ignoring other strategies of children, for instance the use of relational knowledge and estimation (Baroody, 1999b). In a later version of the model, the Adaptive Strategy Choice Model (ASCM; Siegler & Lemaire, 1997; Siegler & Shipley, 1995), information about the efficiency of different strategies is incorporated as well. In this model there are not only associations between problems and answers, but also between certain types of problems and the strategies being used. The ASCM is able to model the shift from nonretrieval procedures to direct memory retrieval in simple arithmetic for both multiplication and addition.

1.5 Addition and multiplication facts

In the research on arithmetic facts, far more attention has been devoted to multiplication than to addition. In the present research, both the learning of simple addition and simple multiplication facts is studied. The participants are children from Grade 1 to Grade 3, with ages between 6 and 9 years. In our research, the simple addition problems used as stimuli have answers up to 10. In contrast, the simple multiplication problems have two factors between 1 and 10, and therefore the answer can be as much as 100. Despite this difference in problem size, there are many reasons to believe that mechanisms involved in storing addition and multiplication facts in memory are similar. For instance, both in addition and in multiplication a problem-size effect and a tie effect is observed. Furthermore, there are similarities in how adults solve simple addition and simple multiplication problems (LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, et al., 1996). There has been some discussion about whether arithmetic facts are stored in a phonological (or sound-based) form. According to the triple-code model of number processing (Dehaene and Cohen, 1995) there are three distinct systems of number representation: a nonverbal quantity system, a verbal system, and a visual system (for Arabic numerals). Dehaene and Cohen argue that both simple multiplications and simple additions are usually stored in rote verbal memory. However, evidence on two brain-damaged patients (Whalen, McCloskey, Lindemann, & Bouton, 2002) suggests that arithmetic facts are not stored and retrieved from memory exclusively in a phonological form. Both patients were often able to retrieve the correct answer to simple arithmetic problems from memory when unable to generate the phonological representation of either the arithmetic problem or the answer to that problem.

Although behavioural responses seem similar, Roussel, Fayol, and Barouillet (2002) found subtle differences in how adults solve simple additions and multiplications. Participants were asked to verify whether a problem presented with an answer was correct or incorrect. In some trials, the operation sign (+ or ×) was presented some time before the operands. This manipulation had a stronger effect in addition problems than in multiplication problems, which suggests that many adults use a counting procedure for additions that coexists with the declarative knowledge stored in the associative network. This counting procedure was activated through the presentation of the + sign, even though the numbers to be added were not visible yet. Furthermore, the problem size effect was stronger in addition than in multiplication. This also points to the use of procedural strategies in addition, because when using a counting procedure, adding large numbers will take more time than adding small numbers. The results of Roussel et al. indicate that although adults' behavioral responses in simple addition and multiplication are similar, direct retrieval from memory is used more often in multiplication than in addition.

Some differences between the storage of addition and multiplication facts may also stem from the educational system that is used. In most countries the use of retrieval is emphasized much more in the case of multiplication than in the case of addition. This is also true for the Netherlands. Nevertheless, Dutch children are not expected to rely exclusively on a retrieval strategy for multiplication. If a child cannot retrieve the answer of a multiplication problem, it is encouraged to use known facts (e.g., $9 \times 8 = [10 \times 8] - 8$ or $4 \times 6 = [2 \times 6] \times 2$) to solve the problem (Treffers, Van den Heuvel-Panhuizen, & Buys, 1999).

1.6 Outline of the following chapters

The aim of the present research is to study automatization in children's mathematics. Chapter 2 focuses on a very basic form of automatization, namely the finding that people in western societies automatically associate small numbers with the left side of space and large numbers with the right side of space. This effect is termed the SNARC effect (spatial-numerical association of response codes). Although much research on the SNARC effect has been conducted over the last years, studies on the SNARC effect in children are scarce. To bridge this gap, we studied the development of the SNARC effect in 7- to 9-year-old children. In Chapter 3, children of the same age group are studied. We look into the development of numerical estimation and try to answer the question of whether the ability to estimate quantity and distance is related to performance in mathematics. Chapter 4 focuses on how to enhance automatization in simple arithmetic. Which type of exercises is most effective for practicing simple addition and multiplication problems? In Chapter 5, transfer effects in children's addition and multiplication are studied. For instance, when children learn $2 + 3$, will they also learn $3 + 2$? And does practicing 7×6 aid children's understanding of 8×6? In Chapter 6, children's improvement on simple addition and multiplication problems is measured. Recent research has shown that individual differences in working memory, counting speed, and rapid naming are related to mathematical ability. Therefore, we aim to find out whether individual differences on cognitive processing tasks can predict children's ability to learn simple arithmetic facts. Finally, in the General Discussion in Chapter 7 the findings in the previous chapters are summarized, the theoretical and educational implications are discussed, and recommendations for further study are presented.

2

Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds

The SNARC (spatial–numerical association of response codes) effect refers to the finding that small numbers facilitate left responses, whereas larger numbers facilitate right responses. The development of this spatial association was studied in 7-, 8-, and 9-year-olds, as well as in adults, using a task where number magnitude was essential to perform the task and another task where number magnitude was irrelevant. When number magnitude was essential, a SNARC effect was found in all age groups. But when number magnitude was irrelevant, a SNARC effect was found only in 9-year-olds and adults. These results are taken to suggest that (a) 7-year-olds represent number magnitudes in a way similar to that of adults and that (b) when perceiving Arabic numerals, children have developed automatic access to magnitude information by around 9 years of age.

Published as: Van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology, 101*, 99-113.

2.1 Introduction

One of the most appealing examples of the close relationship between numbers and space is the SNARC (spatial–numerical association of response codes) effect (Dehaene, Bossini, & Giraux, 1993; Dehaene, Dupoux, & Mehler, 1990). In reaction time experiments, relatively large numbers are responded to faster with a right response than with a left response and relatively small numbers are responded to faster with a left response than with a right response (for an overview of the relations between numbers and space, see Fias & Fischer, 2005; Hubbard, Piazza, Pinel, & Dehaene, 2005). The SNARC effect has previously been found in 9-year-olds (Berch, Foley, Hill, & Ryan, 1999). The current research examines the development of the SNARC effect in more detail. We expect the association between numbers and space to be present in children from a relatively young age, but it may take some years before merely seeing a number automatically activates a spatial representation.

The SNARC effect is usually taken as evidence that humans have an internal representation of numbers as if placed on a horizontal number line, with small numbers on the left side and large numbers on the right side. Dehaene et al. (1993) suspected that the left-to-right direction of this number line is related to the direction of writing in Western cultures. In recent studies, the SNARC effect has been extended to other types of responses besides just manual responses; for example, a SNARC effect can be obtained in a pointing task (Fischer, 2003) or in a task where participants need to answer with eye movements (Fischer, Warlop, Hill, & Fias, 2004; Schwarz & Keus, 2004).

The interpretation that the SNARC effect is caused by the representation of numbers on a left-to-right mental number line, however, has not been unchallenged; for example, Gevers, Reynvoet, and Fias (2003) obtained a SNARC effect on tasks with nonnumerical ordinal stimuli – months of the year and letters. The SNARC effect emerged both when ordinal information was relevant and when it was irrelevant. This suggests that the SNARC effect may be caused by the ordinal property of numbers rather than by numerical quantity. A similar conclusion could be drawn from the study of Ito and Hatta (2004), who tested Japanese participants both on a parity task ("Is the number odd or even?") and on a single-digit magnitude judgment task ("Is the number larger or smaller than 5?"). Although the SNARC effect is commonly found in both of these tasks, a SNARC effect was observed only in the parity task. Ito and Hatta reasoned that the observed SNARC effect in the parity task was due to ordinal information from the odd/even number sequences rather than to quantitative information of numbers.

The study of Ito and Hatta (2004) also challenged the view of the left-to-right direction of the SNARC effect being directly related to the direction of writing in Western cultures (Dehaene et al., 1993). These authors observed a vertical SNARC effect, with large numbers associated with the top choice and small numbers associated with the bottom choice. Because the direction of Japanese writing can be either top to bottom or left to right, a vertical SNARC effect with small numbers associated with the top choice would have been consistent with direction of writing, but instead an effect in the opposite direction was found. Studies with Dutch (Schwarz & Keus, 2004) and Belgian (Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2006) participants showed a vertical SNARC effect in the same direction. Together, these results suggest that people associate large numbers with "up" and small numbers with "down." This association can perhaps be explained by daily life experience; for example, when the quantity of a substance rises (e.g., water in a glass), the level usually goes up. Cultural conventions may strengthen this association; for example, high numbers are typically associated with a top position in mathematical graphs and figures. Nevertheless, although the direction of writing is apparently not the defining factor in the vertical SNARC effect, it may still have a large influence on the horizontal SNARC effect. Strong evidence that the original conclusions of Dehaene et al. (1993) were correct was provided by Zebian (2005), who found that Arabic readers, who write from right to left, show a reversed SNARC effect. Furthermore, in Arabic readers who learned English as a second language, this reversed SNARC effect was weakened. Interestingly, in a study with Chinese readers (Hung, Hung, Tzeng, & Wu, 2008), the orientation of the SNARC effect varied within the same participants, depending on the writing system used in the task; Arabic numbers were mentally aligned with a horizontal left-to-right directionality, whereas Chinese number words were aligned vertically with a top-to-bottom directionality. The different orientations are consistent with the dominant context in which the numerical materials are often encountered, thereby supporting the idea that experience in reading and writing influences the mapping between numbers and space.

The finding of a vertical SNARC effect by Ito and Hatta (2004) led Proctor and Cho (2006) to the conclusion that the metaphor of a horizontal number line to explain the SNARC effect is not necessary at all. They claimed that the SNARC effect in binary classification is a consequence of an asymmetric coding of the stimulus and the response sets instead of being caused by a mental representation of numerical quantity. According to Proctor and Cho, people code the stimulus alternatives and the response alternatives as positive (+) polarity and negative (-) polarity. In the case of the SNARC effect, "large" is coded as + polarity and "small" is coded as - polarity. The + polarity corresponds to "right" in a decision between left and right, and it corresponds to "top" in a decision between top and bottom. Response selection is faster when the polarities coincide than when they do not coincide, and this may explain both the horizontal and vertical SNARC effect. This theory is consistent with evidence that the SNARC effect has its origin in the response selection stage (Keus, Jenks, & Schwarz, 2005; Keus & Schwarz, 2005).

Further evidence against the view that the SNARC effect is directly dependent on a left-to-right number line comes from research showing that the SNARC effect can be influenced by task demands (Bächtold, Baumüller, & Brugger, 1998; Müller & Schwarz, 2007). When participants were instructed to think about numbers as if placed on a ruler, a normal left-to-right SNARC effect was found, but when participants were instructed to imagine numbers on a clock face, the SNARC effect was reversed (Bächtold et al., 1998). The reversed effect is consistent with the clock face representation, which has small numbers on the right side and large numbers on the left side. However, the reversal of the SNARC effect was also associated with longer response times, suggesting that the left-to-right representation was triggered automatically and then interfered with the clock face interpretation. Recently, Müller and Schwarz (2007) found that, in a vertical arrangement of buttons, the SNARC effect could be either hand related or location related, depending on task instruction. Nevertheless, in a horizontal alignment, the location-related number representation was dominant; participants responding with crossed hands exhibited a normal SNARC effect even though task instruction emphasized the mapping of odd/even to the hands. This finding suggests that the (horizontal) SNARC effect is location specific and not hand related. However, in this study, crossed and uncrossed hands were tested within the same session, and this could have influenced the results. In contrast, Wood, Nuerk, and Willmes (2006) failed to find a SNARC effect in a replication of Dehaene et al. (1993) crossed hands study in which participants responded with crossed hands throughout the entire session. Therefore, the finding of a SNARC effect with crossed hands probably holds only under specific conditions.

Overall, it seems clear that the SNARC effect can be influenced by task demands and, therefore, does not rely exclusively on a left-to-right mental number line. At the same time, it is evident that an association between small numbers and "left," and between large numbers and "right," is more natural to people in Western cultures than are other mappings. Given evidence that this bias is culturally defined (Dehaeneet al., 1993; Zebian, 2005), it would be interesting to know the age at which children show signs of an association between small numbers and the left side of space and an association between large numbers and the right side of space. Will it be there as soon as they have learned the meaning of Arabic numerals, or will it take them some years to develop this association? It has been shown that even children who have not yet learned to write will order abstract concepts in the direction that is common to their culture (Tversky, Kugelmass, &Winter, 1991). Furthermore, Opfer and Thompson (2006) observed that, with increasing age, preliterate children develop a directional bias in representing numerical magnitudes. In their study, nearly half of the children between 4.5 and 5.5 years of age placed a new token on the right side of a horizontal row with three tokens and chose the right-most token when asked to take away one of four tokens. This directional bias seems to stem from preferences in counting direction. The association between numbers and space, therefore, may arise quite early in development. However, little research has been devoted to the development of the SNARC effect. The current study was undertaken to increase our understanding of the development of this specific association between numbers and space.

The first study on the development of the SNARC effect (Berch et al., 1999) tested the presence of a SNARC effect in children using a parity task. Children from Grades 2, 3, 4, 6, and 8 made parity judgments (odd/even) of an Arabic numeral by pressing a button left or right. The results showed that as early as Grade 3 (9.2 years of age), children exhibit the SNARC effect. No signs of the SNARC effect were found in children in Grade 2 (7.8 years of age), although Berch et al. (1999) emphasized that the results for this group may be less accurate than those for the other age groups due to a smaller number of participants and highly variable reaction times. However, their results suggest that 9-year-olds (a) associate large numbers with "right" and small numbers with "left" and (b) access magnitude information obligatorily even when irrelevant. This raises the following issue: If indeed children younger than 9 years do not exhibit a SNARC effect in a parity task, then why is that? There are at least three possible explanations: Before 9 years of age, (a) the parity task is too difficult, (b) children do not represent magnitude information in the same way as do adults, and (c) children have no automatic access to magnitude information when perceiving Arabic numerals. In the current study, we sought to clarify this issue by testing 7-, 8-, and 9-year-olds (Grades 1, 2, and 3), as well as adults, under two conditions: (a) in a task where number magnitude is part of the task requirements and (b) in a task where number magnitude is irrelevant (easier than the parity task).

There are a few studies that looked into the development of automaticity in accessing number magnitude. Girelli, Lucangeli, and Butterworth (2000) tested children in Grades 1, 3, and 5, as well as adults, on a numerical Stroop task. Pairs of Arabic numerals were presented in two tasks, and participants needed to choose the digit larger in number magnitude or the digit larger in physical size. Three types of stimuli were presented: (a) congruent pairs where the numerically larger digit was also physically larger (2 6), (b) incongruent pairs where the numerically larger digit was physically smaller (2 6), and (c) neutral pairs. In the physical comparison task (i.e., when numerical values were irrelevant), a size congruity effect appeared; response times were slower for incongruent trials than for neutral ones. This effect was significant from Grade 3 (8.4 years of age) onward and was absent in Grade 1 (6.6 years of age). These findings indicate that children in Grade 1 do not automatically activate information about number magnitude when perceiving numbers. Rubinsten, Henik, Berger, and Shahar-Shalev (2002) refined the results of Girelli et al. (2000) by recruiting children from both the beginning and end of Grade 1. They found a size congruity effect for children at the end of Grade 1 (7.3 years of age) but not for children at the beginning of Grade 1 (6.3 years of age). However, the onset of automatic processing in numbers is probably culturally defined. A recent study on the numerical Stroop task with Chinese children (Zhou et al., 2007) found signs of automatic processing of numerical magnitude even in kindergartners (5.8 years of age). Chinese children have several linguistic and cultural advantages in numerical learning over children in other countries, and this could explain the early onset.

We can conclude from the numerical Stroop tasks that automatization in number processing develops under the influence of growing numerical skills. Therefore, it is possible that Berch et al. (1999) did not find a SNARC effect in children in Grade 2 because at this age children do not automatically activate number magnitude information when confronted with Arabic numerals. Given that Rubinsten et al. (2002) found a size congruity effect in the physical comparison task already in children at the end of Grade 1, there may be another explanation for Berch et al.' (1999) findings; perhaps parity judgment is not a suitable task for young children because "odd" and "even" are rather abstract number properties. A second problem of the parity task is that odd/even may be a semantic property of a digit just like number magnitude. It is possible that magnitude information is activated via a semantic route in the process of retrieving the parity status of a digit; therefore, it would be better to select a task in which number magnitude is completely irrelevant. Such a task should be chosen with care, however, because it has been shown (Fias, Lauwereyns, & Lammertyn, 2001; Lammertyn, Fias, & Lauwereyns, 2002) that participants can ignore irrelevant magnitude information when asked to judge the color of a digit or to identify a shape (circle or square) superimposed on the digit. An elegant design for a task that produces a SNARC effect in adults comes from the work of Fischer, Castel, Dodd, and Pratt (2003), who showed that the SNARC effect truly is an association between numbers and space because merely looking at a number causes a shift in attention to either the right or left visual field. Targets presented on the right side of a screen are detected faster after a large number than after a small number, and vice versa for targets presented on the left side of a screen. We adapted Fischer and colleagues' task for the current research to study the development of automaticity in accessing number magnitude.

To investigate the possibility that children gradually develop an association between numbers and space under the influence of exposure to numbers throughout primary school, we also tested our participants on a simple magnitude judgment task: Is the number smaller or larger than 5? We expected children to have a representation of numbers similar to that of adults before 9 years of age because the distance effect (Moyer & Landauer, 1967) – numerically close numbers are more difficult to compare than are numbers further apart – has been found in children much younger than 9 years; even 5-year-olds show a distance effect in a magnitude comparison task (Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977; Temple & Posner, 1998). The distance effect is consistent with the interpretation that humans represent numbers on a mental number line because close numbers are more difficult to dissociate than are numbers further apart. Therefore, the finding of a distance effect suggests that children use the same representation for numbers as do adults. Nevertheless, recent studies with functional magnetic resonance imaging (fMRI) (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005; Kaufmann et al., 2006) found differences in cerebral activation between children and adults when performing a magnitude comparison task even though behavioral performances were similar.

In the current study, we tested 7-, 8-, and 9-year-olds, as well as adults, both on a task where information about number magnitude was relevant (magnitude judgment task) and on a task where it was irrelevant (detection task [based on Fischer et al., 2003]). The research questions were as follows. First, at what age will children have this SNARC association between small numbers and the left side of space and between large numbers and the right side of space? Second, at what age will children show a SNARC effect when number magnitude is irrelevant? We expected that the SNARC effect would be present from an early age onward in the task where number magnitude must be processed intentionally because studies on the distance effect suggest that children represent number magnitude in a way similar to that of adults. For the magnitude judgment task, we evaluate the distance effect as well as the SNARC effect because the distance effect is often seen as a marker for having automatic access to number magnitude information. In the detection task, where no intentional processing of number magnitude is needed, we expected the onset of the SNARC effect to be later in development because studies on the numerical Stroop task suggest that automatization in number processing develops gradually under the influence of growing numerical skills.

2.2 Method

2.2.1 Participants

Participants were 89 students of two Dutch schools for primary education (Grades 1, 2, and 3) and two groups of adults. The children were tested approximately 3 months before the end of the school year. There were 33 7-year-olds (15 girls and 18 boys), 29 8-year-olds (16 girls and 13 boys), and 27 9-year-olds (16 girls and 11 boys). The mean ages of the children were 7.0 years (*SD* = 0.4), 8.0 years (*SD* = 0.4), and 9.2 years (*SD* = 0.5), respectively. In these groups of children, 21% of the 7-year-olds, 7% of the 8-year-olds, and 15% of the 9-year-olds were left-handed. In the magnitude judgment task, 18 adults (8 women and 10 men) were tested (mean age = 36.7 years, *SD* = 14.2) and 11% were left-handed. In the detection task, 21 adults (12 women and 9 men) were tested (mean age = 31.7 years, *SD* = 9.1) and 38% were left-handed.

2.2.2 Materials and procedure

Magnitude judgment task. The stimulus material consisted of the Arabic numerals 1, 2, 3, 4, 6, 7, 8, and 9. The children were asked to judge whether the presented Arabic numerals were numerically larger or smaller than 5 by pressing left or right. The items were presented in a pseudo-random order; the same number did not appear in consecutive trials, and the correct response could be on the same side at most three times in a row. Each trial started with a black-bordered transparent square (sides 100 pixels, border 2 pixels) on a white background in the middle of the computer screen. The square remained empty for 1 s, and then an Arabic numeral appeared in the center of the square. The stimulus disappeared as soon as the participant pressed a response button or 5 s had elapsed. The interstimulus interval was 1 s. The sequence of the task is presented in Figure 2.1.

There were two experimental blocks with 40 trials each; all numbers were presented five times. There was a break after 24 trials. In experimental Block A, participants were to respond with their left hand to numbers smaller than 5 and with their right hand to numbers larger than 5. In experimental Block B, the key assignments were switched. During the experiment, participants responded with the *a* key (left-hand response) or the *″* key (right-hand response) of the computer's keyboard. The response keys were marked with a white sticker. As a visual aid, two small cards were placed above the two response keys. One card had a picture of a small circle to symbolize numbers smaller than 5, and the other card had a picture of a large circle to symbolize numbers larger than 5[1]. The participants were asked to keep their fingers close to the response buttons. On average, children needed 4 min and adults needed 3 min to finish one block.

A complicating factor with using a magnitude judgment task is that response keys are necessarily intertwined with number magnitude, so it is not possible to test all conditions in the same experimental block. In a pilot experiment, where the first block was positioned at the beginning and the second block was positioned at the end of a 30-min session, the results indicated that children in Grade 1 slowed down in the second block. This could be due either to fatigue or to confusion about the change in response keys. To prevent these effects in the current study, we tested children in two short sessions of 10 min each with 1 week in between.

Detection task. The design of this task was adapted from Fischer et al. (2003) (see Figure 2.1 for the task sequence). Trials started with a black point (width = 5 pixels) between two boxes (sides = 75 pixels, border = 2 pixels). After 400 ms, the point was replaced by a number from 1 to 9, which was replaced by a point again after 400 ms. Then a "short" delay (400 ms) or a "long" delay (750 ms) elapsed. The reason for a variable delay was to discourage participants from guessing the timing of the sequence. Both delays are within the range where Fischer and colleagues found a SNARC effect. After the delay, one of the two boxes was colored gray and participants needed to press the space bar to indicate that they had seen the target. The boxes disappeared as soon as the space bar was pressed or 5 s had elapsed, and an empty screen was presented for 1 s between all trials. Participants responded with their preferred hand and were asked to keep their hand close to the response button. The instruction emphasized that participants should look at the point in the middle of the screen and pay attention to the presented numbers.

There were two experimental blocks that differed only in the order of the trials. Each block had 36 trials, and there was a break after 18 trials. The numbers 1 to 9 were presented two times together with a left target and two times together with a

---

[1] The cards with the symbols partly covered the number line on the keyboard, so it was unlikely that children could glance at the number line while doing the task. Furthermore, none of the children mentioned the number line on the keyboard, and we did not observe them paying attention to these keys.

right target. Long delays (750 ms) and short delays (400 ms) were distributed evenly between the conditions. The order of the trials was pseudo-random; the same number did not appear in consecutive trials, and the correct response could be on the same side at most three times in a row. Furthermore, the delay was at most three times "short" or three times "long" in consecutive trials. On average, children needed 3.0 min and adults needed 2.5 min to finish one block.

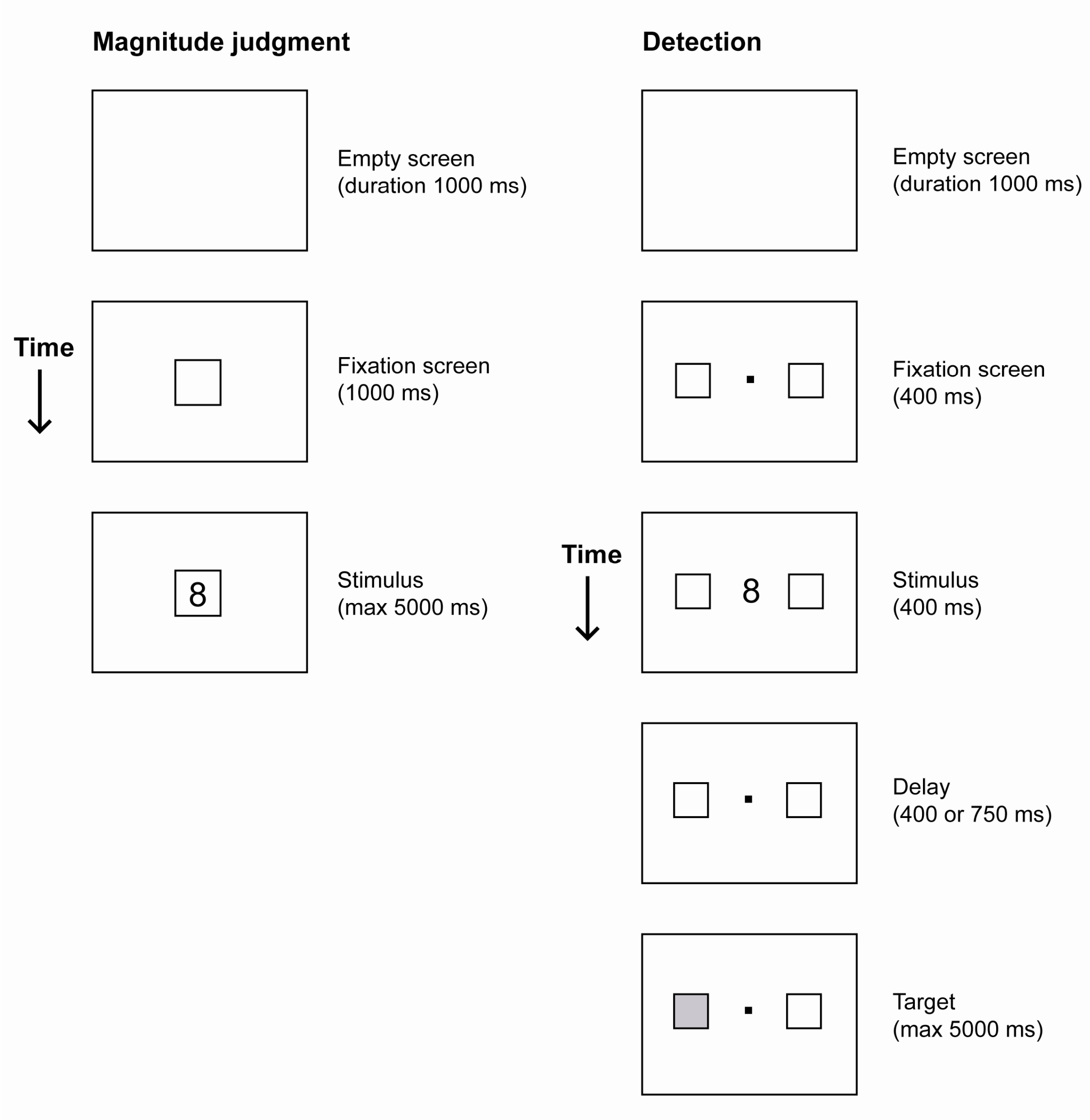

Figure 2.1. Sequence of a trial in the magnitude judgment task (left) and the detection task (right).

Procedure. The children were tested in two sessions of 10 min each with 1 week in between. In both sessions, they performed one experimental block from the magnitude task and one experimental block from the detection task. The sessions were kept short to prevent fatigue, and the period in between was inserted to prevent confusion about response buttons in the magnitude task. Half of the children started with the magnitude task in both sessions, and the other half started with the detection task. In both orders, roughly an equal number of children started with Block A in the magnitude task.

The tasks were conducted with two different groups of adults; the experimental blocks of the tasks were presented immediately after each other. In the magnitude task, half of the adults started with Block A. The tasks were conducted on a laptop with a 15-inch screen and a display resolution of 800 × 600 pixels. The stimuli were presented in Arial font size 48. An experimental block was preceded by at least 10 practice items; this number could be increased to 15 or 20 items if the experimenter deemed it necessary based on the participant’s performance.

2.2.3 Analysis

Magnitude judgment task. One single reaction time (RT) below 200 ms was removed first, and then for each block incorrect trials and outliers were removed. An RT was considered an outlier if it differed by more than 2.5 standard deviations from the mean RT of a participant in a block. If more than 20% of the trials (≥ 9 of 40) needed to be removed, the participant was excluded from the analysis. Furthermore, when more than three of five trials were missing in a specific combination of a numeral and a response side, the participant was excluded as well. Of the 89 participating children, the data from 10 (6 7-year-olds and 4 8-year-olds) were not analyzed in the magnitude task because these children had too many trials rejected. For the remaining 79 children, a total of 7.5% of the trials were removed, whereas the adults had 4.1% missing trials.

Detection task. RTs below 200 ms were removed from the data of the children (3.4%), and RTs below 150 ms were removed from the data of the adults (2.0%). The criterion was higher for the children than for the adults because children’s responses are generally slower. Outliers were calculated over the total number of trials. When an RT differed by more than 2.5 standard deviations from the mean RT of a participant, it was removed. All 89 participants were included in the data set because none of them missed more than 20% of the trials or missed more than two of four in a specific combination of a numeral and a response side. The children had 6.0% missing trials, whereas the adults had 3.9% missing trials.

2.3 Results

2.3.1 Magnitude judgment task

The data were analyzed using linear regression analysis as proposed by Fias, Brysbaert, Geypens, and d’Ydewalle (1996). In a first step, for each participant the mean RT was computed for each number separately for left and right responses. Then difference scores (dRTs) were computed by subtracting the mean RT for left-hand responses from the mean RT for right-hand responses. The finding of a SNARC effect would be visible in the negative slope of a linear regression line when dRT is plotted as a function of number magnitude because right-hand responses are faster with larger numbers and left-hand responses are faster with smaller numbers. However, in this task the order of the blocks should be taken into account as well because the two experimental blocks had different assignments of response keys. Therefore, an extra variable for the order of the blocks was added to the regression equation; numbers that were first responded to with the left hand were coded as −0.5, and numbers that were first responded to with the right hand were coded as +0.5.

There were not sufficient trials per participant (two to five in every combination of a numeral and response side) to reliably calculate a linear regression line for every single participant; therefore, a general regression analysis was performed for each age group. The dependent variable was dRT. In a first step, order was added as a predictor to control for effects of the order of the blocks. In a second step, magnitude was added as a predictor. See Table 2.1 for the results of the regression analyses. The expected negative slope between number magnitude and dRT deviated significantly from zero in each age group. Regression lines are presented in Figure 2.2. In 7- and 8-year-olds there was only an effect of magnitude, but in 9-year-olds there were effects of both magnitude and order. In Figure 2.3, the mean difference scores in 9-year-olds for numbers smaller than 5 and numbers larger than 5 are presented separately for both orders. The graph shows that children who started with Block A (small–left, large–right) had a smaller SNARC effect than did children who started with Block B. This implies that the 9-year-olds responded faster the second time they were tested, perhaps because they remembered the task from the first time, leading to increased difference scores for children who started with Block B and reduced difference scores for children who started with Block A. However, a SNARC effect is clearly present for both orders, and magnitude was indeed highly significant in the regression analysis despite the effect of order. Adults showed no effect of order, only an effect of magnitude.

Table 2.1. Results of regression analyses of the magnitude judgment task.

Age group		Predictor	B	Beta	ΔR^2	*p*
7-year-olds	Step 1	Order	-22.6	-.03	.00	.627
	Step 2	Magnitude	-47.5	-.36	.13	.000
8-year-olds	Step 1	Order	62.1	.11	.02	.118
	Step 2	Magnitude	-19.6	-.19	.04	.007
9-year-olds	Step 1	Order	-80.9	-.15	.01	.019
	Step 2	Magnitude	-33.6	-.35	.12	.000
Adults	Step 1	Order	-3.7	-.02	.00	.759
	Step 2	Magnitude	-10.6	-.38	.14	.000

Note. The order of the blocks and number magnitude were added in two steps as predictors for the difference in RT between right- and left-hand responses.

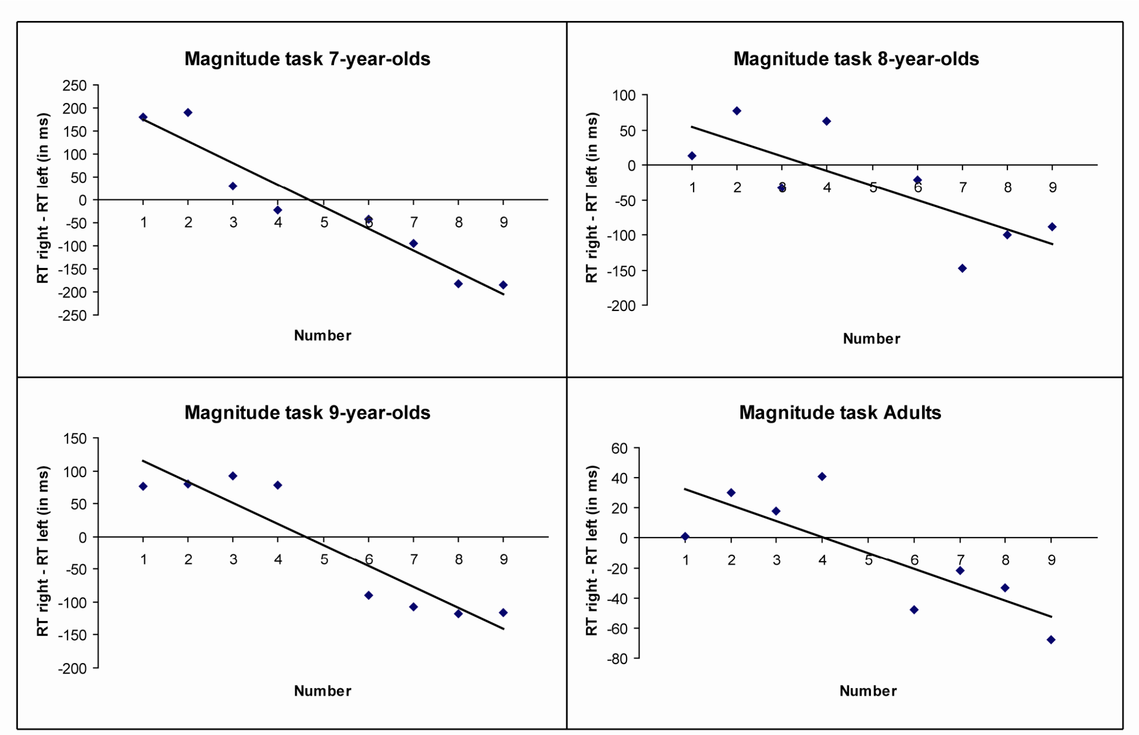

Figure 2.2. Observed data and regression lines for the magnitude judgment task plotted separately for each age group. The difference in RT between right- and left-hand responses (in ms) is regressed on number magnitude.

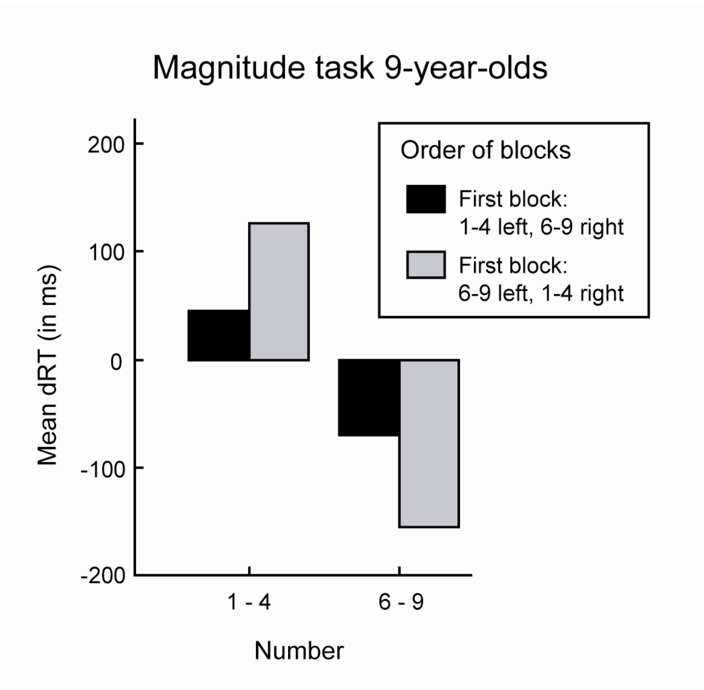

Figure 2.3. Effect of order on the SNARC effect in 9-year-olds. Mean dRT for small numbers (1-4) and for large numbers (6-9) are shown separately for children who started with Block A (black bars) and for children who started with Block B (gray bars).

The data points in Figure 2.2 suggest that the SNARC effect in the magnitude task becomes more categorical with increasing age. Furthermore, the slope of the regression line becomes less steep with increasing age, as can be concluded from the decreasing weight of the magnitude variable in the different equations. This effect is presumably caused by faster RTs in older participants (Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006). See Table 2.2 for the mean RTs. The difference in weight between 7-year-olds and 8-year-olds was significant because in a separate regression analysis for the two age groups a significant interaction effect between age and magnitude was found, $t(3, 412) = -2.40$, $p < .05$. The weight of magnitude in 9-year-olds could not be compared with the other groups because of the interfering effect of order. Accuracy data were compared between the blocks, and only 7-year-olds showed a significant difference in errors, $t(26) = -2.57$, $p < .05$. They made more errors in Block B (6.4%) than in Block A (3.9%), consistent with the SNARC effect. The order of the two tasks in a session did not influence the results of the magnitude judgment task.

Table 2.2. Mean reaction times (in ms) of the four age groups in the magnitude judgment task and the detection task.

	Magnitude		Detection		Detection short		Detection long	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
7-year-olds	1038	281	512	182	530	177	495	185
8-year-olds	877	190	472	131	492	132	452	128
9-year-olds	826	197	443	152	460	147	426	155
Adults	495	74	344	70	354	66	333	72

Note. For the detection task, reaction times are given separately for short and long intervals.

To evaluate the distance effect, we computed a linear regression equation for each participant with the logarithm (base 10) of the distance (1, 2, 3, 4) from the reference number (5) as the predictor variable for RT. After this, *t* tests were performed to test whether the regression weights deviated significantly from zero. The mean equations for the groups are listed in Table 2.3.

Table 2.3. Evaluation of distance effect in the magnitude judgment task.

Age group	Mean equation	*t* test	*p*
7-year-olds	RT = 1105.5 – 194.4(distance)	$t(26) = -3.22$	.003
8-year-olds	RT = 923.8 – 135.3(distance)	$t(24) = -3.30$	.003
9-year-olds	RT = 871.0 – 130.6(distance)	$t(26) = -4.46$	.000
Adults	RT = 524.9 – 86.4(distance)	$t(17) = -9.21$	.000

Note. Individual linear regression lines were calculated with the logarithm of the distance as the predictor for mean RT. In each age group, the regression weights were tested against zero with a *t* test. The second column shows the mean equation for distance over the whole group.

In each age group, the regression weights for distance differ significantly from zero. The negative correlation between numerical distance and RT indicates that the expected distance effect was present in each age group. Moreover, consistent with previous studies (Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977), the slope of the regression line becomes less steep with increasing age. The slope of the adults is significantly different from the slopes of the 7-year-olds, $t(43) = -1.45$, $p < .01$, the 8-year-olds, $t(41) = -1.00$, $p < .05$, and the 9-year-olds, $t(43) = -1.20$, $p < .01$. Furthermore, the slope of the 9-year-olds differs significantly from the slope of the 7-year-olds, $t(52) = -0.95$, $p < .05$.

2.3.2 Detection task

In the detection task, difference scores were computed for each number by subtracting the mean RT for left-side targets from the mean RT for right-side targets. There was no additional variable for the order of the blocks because the conditions in the detection task were evenly distributed across trials. From Table 2.2, it can be seen that RTs were faster for long intervals than for short intervals. There were not sufficient trials to statistically evaluate the effect of magnitude on the difference scores separately for short and long intervals, but inspection of the data in 9-year-olds and adults showed the same downward trend for short and long intervals. Regression lines were calculated separately for the two intervals for both 9-year-olds and adults. First, an average RT for left-side targets and for right-side targets was computed over the whole group. After that, difference scores were obtained by subtracting the RT for left-side targets from the RT for right-side targets, leading to nine data points. Finally, a regression equation was calculated. The effect of magnitude did not reach significance in any of the linear regressions, although there was a trend for an effect on the short intervals in 9-year-olds, $t(1, 7) = -2.26$, $p = .059$. Nevertheless, the slope was negative in all regression equations: 9-year-olds, short intervals, dRT = 27.0−8.4 (magnitude); 9-year-olds, long intervals, dRT = 21.1−2.6 (magnitude); adults, short intervals, dRT = 23.7−4.5 (magnitude); adults, long intervals, dRT = 13.1−2.7 (magnitude). Therefore, we considered it safe to collapse the data over interval conditions.

Table 2.4. Results of regression analyses of the detection task with number magnitude as the predictor for the difference in RT between right- and left-hand responses.

Age group	Predictor	B	Beta	R^2	p
7-year-olds	Magnitude	-2.3	-.06	.00	.341
8-year-olds	Magnitude	-0.8	-.03	.00	.672
9-year-olds	Magnitude	-5.2	-.14	.02	.025
Adults	Magnitude	-3.8	-.23	.05	.001

The results of the regression analyses are presented in Table 2.4. Magnitude did not have a significant effect on 7- and 8-year-olds, but there was a negative slope between number magnitude and dRT in 9-year-olds and adults. The negative slope implies a SNARC-like effect in these two age groups. Observed power ($1 - \beta$) was .62 in 9-year-olds and .90 in adults. The graphs showing the data points and regression lines for the groups are presented in Figure 2.4. The regression lines of 7- and 8-year-olds, which were not significant, are presented as dotted lines. The order of the two tasks in a session did not influence the results of the detection task.

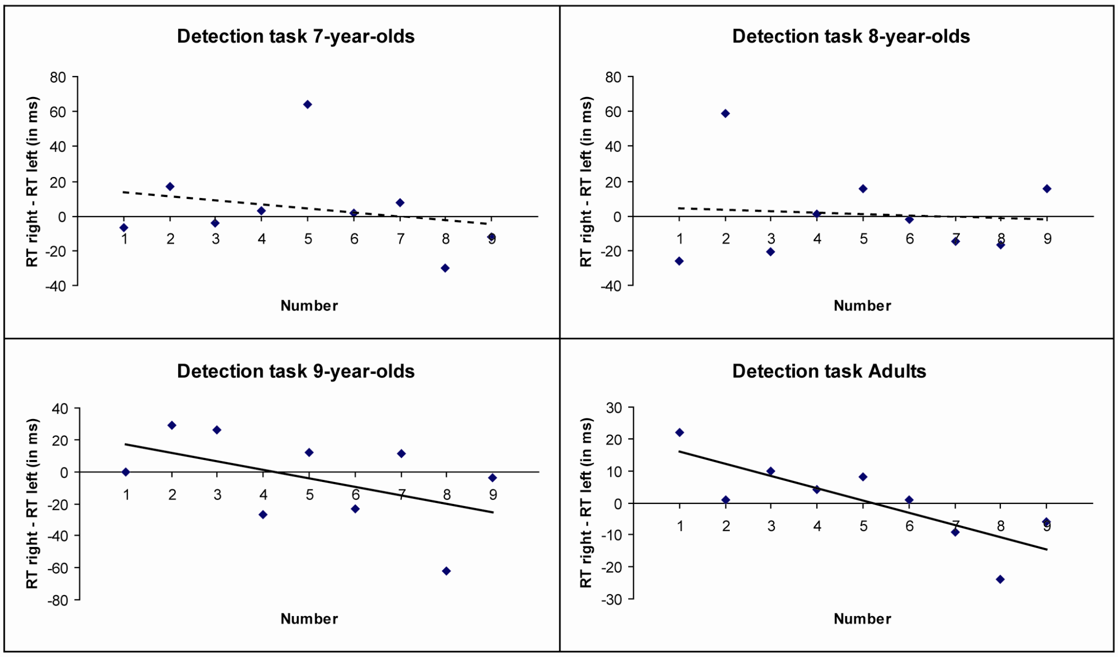

Figure 2.4. Observed data and regression lines for the detection task displayed separately for each age group. The difference in RT between right- and left-side targets (in ms) is regressed on number magnitude. Linear regression is significant in 9-year-olds and adults but not in 7- and 8-year-olds (dotted line).

2.4 Discussion

The current research was undertaken to study the development of the SNARC effect. Participants from different age groups performed a magnitude judgment task and a detection task. In the magnitude task, number magnitude was part of the task requirements; participants judged whether a number was numerically smaller or larger than 5 by pressing a button left or right. In this task, a SNARC effect was found in 7-, 8-, and 9-year-olds as well as in adults; small numbers were responded to faster with the left hand than with the right hand, and vice versa for larger numbers. Furthermore, the distance effect was present in all age groups; that is, RTs were slower for numbers close to the standard than for numbers further away from the standard. In the detection task, number magnitude was irrelevant; after the presentation of a number, participants pressed a button as soon as they noticed the target. In this task, the SNARC effect was absent in 7- and 8-year-olds; only 9-year-olds and adults showed a SNARC-like effect. The results strongly suggest that (a) children have an association between small numbers and “left” and between large numbers and “right” when they are as young as 7 years and that (b) children have automatic access to magnitude information when perceiving Arabic numerals from 9 years of age.

The current study contributes to the understanding of the results from Berch et al. (1999). In their study, using a parity task, a SNARC effect was found not before 9 years of age. Our results on the detection task suggest that the findings of Berch and colleagues cannot be attributed to difficulty in making parity judgments. In line with what they found, our results corroborate the hypothesis that children will not automatically show a SNARC effect before they are approximately 9 years of age. This later manifestation of the SNARC effect is not necessarily caused by a number representation in younger children different from that in adults. Instead, it seems to be related to the fact that younger children do not automatically activate semantic information about number magnitude. An association between numbers and space, therefore, will appear only in tasks that directly involve number magnitude. The magnitude task used in our study to determine the presence of the SNARC association indeed revealed an effect at earlier ages. At a later stage of development, when children will have had more experience with numbers, number magnitude information is automatically activated on seeing a number. At this point in development, a SNARC effect can be found in other tasks, such as parity and detection tasks, even though number magnitude is not addressed directly.

On the other hand, our results about automatic number processing at approximately 9 years of age seem contradictory to the results of Rubinsten et al. (2002). Using a numerical Stroop paradigm, they found evidence for automatic access to magnitude information at the end of Grade 1 (7.3 years of age) in the physical comparison task. Early signs of automatization in number processing may well depend on the specific task used. A major difference between the detection task and the numerical Stroop task is the basis of responding. In the detection task participants simply responded to the appearance of a gray color after a number was presented, whereas in the physical comparison task numbers were present and played a role in the comparison. Higher degrees of automatization in number processing may be necessary before automatization is demonstrated in a detection task, whereas a Stroop task may detect lower degrees of automatization. In addition, different studies on the onset of automatic number processing may be difficult to compare because of age differences at the beginning of schooling and sociocultural differences in the emphasis on knowledge about numbers during the preschool or kindergarten period. The contrast between the finding of a size congruity effect in Chinese kindergartners (5.8 years of age) (Zhou et al., 2007) and no effect in children beginning Grade 1 in Israel (6.3 years of age) (Rubinsten et al., 2002) or in children in Grade 1 in Italy (6.6 years of age) (Girelli et al., 2000) may be explained by such differences.

An alternative explanation for the absence of the SNARC effect in the lowest age groups on the detection task is that the 7- and 8-year-olds did not give the presented numbers as much attention as did the 9-year-olds. However, in the instructions it was underlined that participants should look at the numbers, and in the actual experimental situation the presented numbers were hard to ignore.

Therefore, it seems unlikely that the younger children did not process the presented numbers. Overall, the finding of a SNARC effect in 9-year-olds in the detection task suggests that this design is suitable for research on children. This is an important finding for future research on the development of number processing.

The current results of the magnitude judgment task seem to show a gradual change from continuous to categorical data over age. The adult pattern is commonly found in magnitude comparison tasks (Gevers et al., 2006). One reason for a categorical effect might be that for judging number magnitude only a rough categorization, such as “smaller than 5” or “larger than 5,” is needed. However, Gevers et al. (2006) explained the categorical effect from an interaction between the SNARC effect and the distance effect. The distance effect causes RTs for numbers close to the standard (4 and 6) to be much slower than those for numbers further away. Because the SNARC effect becomes stronger as RTs are slower, the SNARC effect will have a greater effect on the numbers near the standard. Therefore, the difference between the left response and the right response increases, and this pushes up the difference score for 4 and pushes down the difference score for 6, leading to a categorical shape. In the current study, the results seem to show that the SNARC effect in the magnitude judgment task becomes more categorical with increasing age. The reason for this cannot be that the distance effect is stronger in higher age groups, and thus has a larger effect, because the distance effect actually becomes smaller with higher age, as was found in previous research (Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977). Thus, the finding of a change over ages from continuous to categorical data in the magnitude judgment task cannot easily be explained by an interaction between the SNARC effect and the distance effect.

What else might explain the change from continuous to categorical patterns? It is possible that younger children process numbers in a way different from that of adults, for example, because children compare numbers using an algorithmic procedure, whereas adults indeed use rough categorizations as “small” and “large.” The current experimental design does not allow us to determine whether children and adults used the same brain regions during the task, but we know from other studies (Ansari et al., 2005; Kaufmann et al., 2006) that primarily frontal cortical areas are engaged in children during magnitude comparison, whereas parietal regions are activated in adults. This suggests that, even when responses are highly similar at a behavioral level, children and adults process numbers in different ways. Therefore, it is possible that the youngest children used a different procedure to retrieve number magnitude in the magnitude judgment task.

A distance effect in the magnitude judgment task was found in all age groups. From this finding, we infer that participants from each age group used a quantitative strategy to solve the task. Assuming that the distance effect is a reliable marker for having access to number magnitude information, we reason that in all age groups access to number magnitude was established. Finding both a SNARC effect and a distance effect in the children suggests that children represent numbers in the same way as do adults; close numbers are harder to discern than are numbers further away, and larger numbers are associated with the right side of space. The SNARC effect and the distance effect are often interpreted as evidence that humans represent numbers on a mental number line. However, Ito and Hatta (2004) surprisingly observed no SNARC effect in adults on a magnitude judgment task, although there was a distance effect. Therefore, the two effects might stem from different origins. Furthermore, in studies on the numerical Stroop task, a distance effect was often not observed in the physical comparison task even when the size congruity effect was present (Girelli et al., 2000; Rubinsten et al., 2002; Tzelgov, Meyer, & Henik, 1992). Rubinsten et al. (2002) proposed that the distance effect might be considered a product of the intentional processing of number magnitude and does not arise when knowledge about number magnitude is not triggered by the task design (see also Noël, Rousselle, & Mussolin, 2005; Tzelgov & Ganor-Stern, 2005). Therefore, finding both a SNARC effect and a distance effect does not necessarily mean that children represent numbers on a mental number line. Alternative explanations, such as an asymmetric coding of the stimulus and the response sets (Proctor & Cho, 2006), are also consistent with the current data, although the continuous shape of the SNARC effect in the 7- and 8-year-olds suggests an algorithm-based process in at least the youngest children. To clarify this issue, more research is needed. For instance, it would be interesting to use the design of Müller and Schwarz (2007) with children to find out whether a vertical SNARC effect is present and also whether horizontal and vertical SNARC effects can be influenced by task instruction. If the SNARC effect in children is easily influenced by task instruction, an asymmetric coding of the stimulus and the response sets would be a more appropriate explanation for the finding of a SNARC effect in the magnitude judgment task than would the representation of numbers on a mental number line.

The results of the magnitude judgment task demonstrate that children at the end of the first year of formal education, or at least the majority of these children, have formed an association between small numbers and the left side of space and between large numbers and the right side of space. However, the study does not give a detailed answer to the question of the age at which the spatial association arises because children in kindergarten or at the beginning of Grade 1 were not included. Of course, children will be able to show a SNARC effect only if they are familiar with the meaning of Arabic numerals, so this would complicate testing in kindergarten classes, at least in The Netherlands. Clarifying the issue of whether a SNARC effect will emerge as soon as children have mastered the meaning of Arabic numerals, therefore, would require a careful developmental study of younger children with a focus on individual differences in numerical skills. Nevertheless, based on the results of the current study, it does not seem conceivable that children are slow in developing a SNARC association. In contrast, it appears that only little experience with numbers is sufficient to associate small numbers with ‘”left” and large numbers with “right.”

Unfortunately, the design of the current study did not provide an opportunity to look at individual differences among children because that would require substantially more trials to calculate regression lines for each individual participant. Future research could, for example, examine the number of children who show a SNARC effect in each grade because the SNARC association might not be established in every child at the same period of time during development. Furthermore, it is unknown whether differences exist between individuals in the strength of the SNARC effect. Recent studies reported interesting findings about the links among basic number processing, visuospatial abilities, and mathematical skills (Bachot, Gevers, Fias, & Roeyers, 2005; Rouselle & Noël, 2007; Rubinsten & Henik, 2005). Studies with more trials than the current study could possibly measure for every child whether there is a large SNARC effect, a small SNARC effect, or no effect at all. These individual differences may be linked with visuospatial abilities and mathematical skills.

To conclude, this study with 7-, 8-, and 9-year-olds shows that from the age of 7 years onward, children represent information about number magnitude in a similar way as do adults, and that from the age of 9 years onward, children have direct access to this number magnitude information when perceiving Arabic numerals.

2.5 Acknowledgments

The authors thank Willeke Feenstra for her help in collecting the data for this study.

3

Development of numerical estimation and its relationship with mathematics in Grade 1 to 3

This study examines numerical estimation in children of Grade 1, 2, and 3 (mean age 7.2, 8.2, and 9.2 years). Four domains of estimation were tested: number line, numerosity, length, and area. Principal components analysis revealed two different factors: 1) estimation of area and length, 2) number line and numerosity estimation. The two factors corresponded to different developmental trajectories over grades. Mathematics achievement scores correlated with 3 of the 4 estimation domains in Grade 1, but the relation between estimation and mathematics disappeared in higher grades. Although there are some methodological limitations, it is tentatively concluded that estimation skill and general ability in mathematics are more intertwined in younger children than in older children.

3.1 Introduction

Although estimation skills are clearly useful in everyday life, for instance when judging the distance to your house and how much time it will take you to get there, children usually perform poorly on estimation problems (Crites, 1992; Montague & Van Garderen, 2003). Recent studies suggest there may be a link between estimation skill and general mathematical ability (Booth & Siegler, 2006; Dowker, 2003). The present study aims to investigate the development of four different categories of numerical estimation in children of Grade 1, 2, and 3, using an estimation test with the following problem types: number line estimation, numerosity estimation, and length and area estimation using a given reference. Research questions are: How do estimation skills develop over grades? Are the four domains of estimation interrelated? And finally, is there a relationship between estimation skill and mathematical ability?

The focus in the present study is on numerical estimation, which refers to translating inexact quantities to exact numbers, or vice versa. The most extensive study on numerical estimation in children comes from Booth and Siegler (2006). They took the developmental trajectory of number line estimation – translating the magnitude of a number into a spatial position on a line – as a starting point to study the development of other types of numerical estimation. There has been a substantial amount of research on the development of skill in number line estimation. Accuracy on number line estimates improves considerably during elementary school (Petitto, 1990; Siegler & Booth, 2004; Siegler & Opfer, 2003). This improvement is caused on the one hand by better strategy use: older children use anchor points like the midpoint or the quarter values, or they count by multiples of 5 or 10 to find the answer (Petitto, 1990; Siegler & Opfer, 2003). On the other hand, accuracy is raised by an internal representation of numerical magnitude that shifts from logarithmic to linear. This shift has been observed for young children (6-8 years) on the 0-100 number line (Siegler & Booth, 2004), as well as for older children (8-12 years) on the 0-1000 number line (Siegler & Opfer, 2003).

Booth and Siegler (2006) hypothesized that the shift from a logarithmic towards a linear representation of magnitudes is probably not only important for the development of number line estimation, but also for other types of numerical estimation. They tested Grade 2 and Grade 4 students (mean age 7.8 and 9.9 years) on number line, numerosity, and length estimation. The magnitudes to be estimated were in the 0-1000 domain. Results showed that performance on the estimation tasks was higher in Grade 4 compared to Grade 2. Differences in performance, both between grades and within grades, could be explained effectively from an increasing reliance on a linear magnitude representation. In addition, individual scores on the three tasks correlated substantially with each other, both in Grade 2 and in Grade 4. These results suggest a general cognitive development underlying different types of numerical estimation. Moreover, the scores on the three estimation tasks were substantially correlated to scores on a standard mathematics achievement test, thus indicating a relation between skill in numerical estimation and mathematical ability.

Apart from numerical estimation, Booth and Siegler also tested proficiency in computational estimation. Computational estimation refers to giving approximate answers to arithmetic problems without doing an exact calculation. This skill appears to be highly associated with skill in exact computation (Dowker, 2003; Lefevre, Greenham, & Waheed, 1993). The results of Booth and Siegler support this view; although performance on computational estimation was correlated to performance on the other estimation tasks, the correlations disappeared when the mathematics achievement scores were partialled out from the analysis. This finding suggests that computational estimation is more closely linked to mathematical ability than to other types of numerical estimation.

In a study with adults, Hogan and Brezinski (2003) tested computational estimation, numerosity estimation, and estimation of measurement (length, weight, volume, and time). Their aim was to determine the relations among the estimation abilities and to find out how they fit into general mathematical ability. For this purpose, subjects performed two mathematical tasks: one with basic arithmetic operations and one with mathematical word problems. The scores on the estimations tasks and the mathematical tasks were entered in a principal components analysis. This is a technique for finding underlying (or latent) variables. It can provide an understanding of how different variables are connected to each other. Two components were identified, the first one aligned computational estimation with the two mathematical tasks, and the second component unified numerosity and measurement estimation. These results indicate that numerosity and measurement estimation form a unique estimation skill, apart from computational estimation and general mathematical ability.

Based on the studies of Booth and Siegler (2006) and Hogan and Brezinski (2003), computational estimation appears to be tightly connected to mathematical ability, whereas number line, numerosity and measurement estimation form another coherent category. Nevertheless, Booth and Siegler also found relations between mathematical ability and skill in numerical estimation, suggesting that good mathematical skills are advantageous for numerical estimation, or vice versa. This does not mean, however, that all types of numerical estimation are related to mathematical ability in the same way. Pike & Forrester (1997) tested two types of measurement estimation, length and area, in children from 6 to 11 years old. They found that scores on area estimation were related to mathematical skills, whereas scores on length estimation were not. Intuitively, estimating area seems similar to estimating length, and therefore these findings may come as a surprise. On the other hand, good multiplication skills are highly beneficial for estimating the area of rectangular figures (at least if children realize that there is a relation between the two sides of a rectangle and its area), whereas children cannot benefit from multiplication skills when estimating length. Therefore, the results of Pike and Forrester (1997) raise the question of whether length and area rely on the same competences.

In the present study, four different categories of numerical estimation will be investigated: number line, numerosity, length, and area. Children from Grades 1, 2, and 3 are included to obtain a clear view on the early development of estimation skill. The results of the estimation test will first be analyzed with a reliability analysis, in order to evaluate whether the estimation task provides a good measure of skill in estimation. After that, the relationships among the four estimation tasks will be studied using a principal components analysis, because this seems a fruitful approach for detecting associations between different variables. Furthermore, the correlations between individual estimation scores and scores for general mathematical ability will be calculated. Based on the results of previous studies (Booth & Siegler, 2006; Hogan & Brezinski, 2003) it is anticipated that the four estimation tasks are related to each other and follow the same developmental trend. Concerning the relationship between estimation skill and general mathematical abilities, positive correlations are expected between scores for estimation and mathematics.

3.2 Method

3.2.1 Participants

The participants were 237 students from five Dutch schools for primary education and involved children from Grade 1, 2, and 3 in the second half of the school year. The number of participants was 83 in Grade 1 (34 girls, 49 boys), 78 in Grade 2 (47 girls, 31 boys), and 76 in Grade 3 (41 girls, 35 boys). Mean age of the children was 7.2 years, 8.2 years, and 9.2 years, respectively.

3.2.2 Mathematical ability

As an index of mathematical skills the TTR (de Vos, 1992) and the most recent administered mathematics test of the pupil monitoring system of CITO (2002) were used, both of which are commonly used tests in the Netherlands. In the TTR a series of simple additions and subtractions is listed and children are asked to answer as much items as possible in 1 minute. The TTR was administered prior to the estimation test; children performed 1 minute of additions and 1 minute of subtractions. The pupil monitoring system tests a broader range of arithmetic skills: it includes subjects like counting, ordering numbers, breaking up numbers into smaller numbers, measurement, time, and money. Most of the problems are provided within a realistic context. The CITO tests are administered twice in every school year; the most recent test scores were obtained 2-3 months before this experiment was carried out.

The score on the TTR consisted of the number of items answered correctly within the two minutes of the test. Because the same test was used in all three grades, scores increased over grades. Mean scores were 17.9 (*SD* = 5.8), 33.3 (*SD* = 7.8), and 39.9 (*SD* = 8.3), respectively. The mathematics test of CITO was scored on a scale that provides an index of the arithmetic skills that the student has mastered. The same general scale was used in all three grades, that is, although the children made different tests, the raw test scores were transformed into a score on the same general scale of skill level. Therefore, scores increased over grades: mean scores were 41.5 (SD = 15.0), 65.7 (SD = 8.6), and 79.6 (SD = 11.4), respectively. As a check for reliability, the correlations between the scores on the two arithmetic tests were calculated. A significant positive correlation (p < .001) was found in each grade: r was .61, .64, and .49, respectively. The lower correlation in Grade 3 might be caused by the introduction of new elements (measurement, time, and money) into the CITO test.

3.2.3 Material

The estimation test consisted of 48 problems. There were four different estimation tasks with 12 problems each. The order of the tasks (and also the order of the problems within a task) was the same for all students: 1) Number line, indicating the location of a number on the number line from 0 to 60, 2) Area, estimating how many small rectangles would fit into a larger rectangle, 3) Length, estimating how many small lines would fit into a larger line, and 4) Numerosity, estimating the number of dots. Examples of the estimation problems are provided in Figure 3.1. Prior to the present study, the tasks had been pretested in a pilot experiment in which children were tested individually.

Number line. A 12 cm number line was printed across the middle of the page, with 0 on the left side, 60 on the right side, and no numbers in between. A small flag with a variable number was printed on the right side of the page, above the number line. The children were asked to find the right location for the flag on the number line and to mark this spot with a vertical line. The numbers on the flag were 6, 10, 15, 19, 24, 28, 33, 37, 41, 46, 52, and 59. The problems were presented in random order.

Area. A picture of a stamp was presented next to a rectangle. The children were asked to indicate how many of these stamps would fit into the rectangle without leaving space between the stamps. Correct answers were 6 (2×3), 8 (2×4), 9 (3×3), 12 (3×4), 15 (5×3), 16 (2×8), 20 (5×4), 21 (3×7), 24 (4×6), 28 (4×7), 30 (5×6), and 32 (4×8) stamps. The problems were presented in random order. The size of the reference stamp was the same on every page.

Length. A horizontal line of variable length was shown above a kangaroo that made a single jump. The length of the jump was indicated by a smaller horizontal line. The children were asked to estimate how many jumps the kangaroo would need to make if he wanted to go from left to right over the horizontal line. Correct answers were 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, and 15 jumps. The problems were presented in random order. The length of the reference jump was the same on every page.

Numerosity. The children estimated how many marbles were printed on a page, without counting. The marbles were irregularly spaced, but not far from each other. All marbles had the same size. Area and density were not controlled for. Correct answers were 11, 14, 16, 22, 28, 34, 47, 49, 52, 56, 63, and 75. The problems were presented in random order.

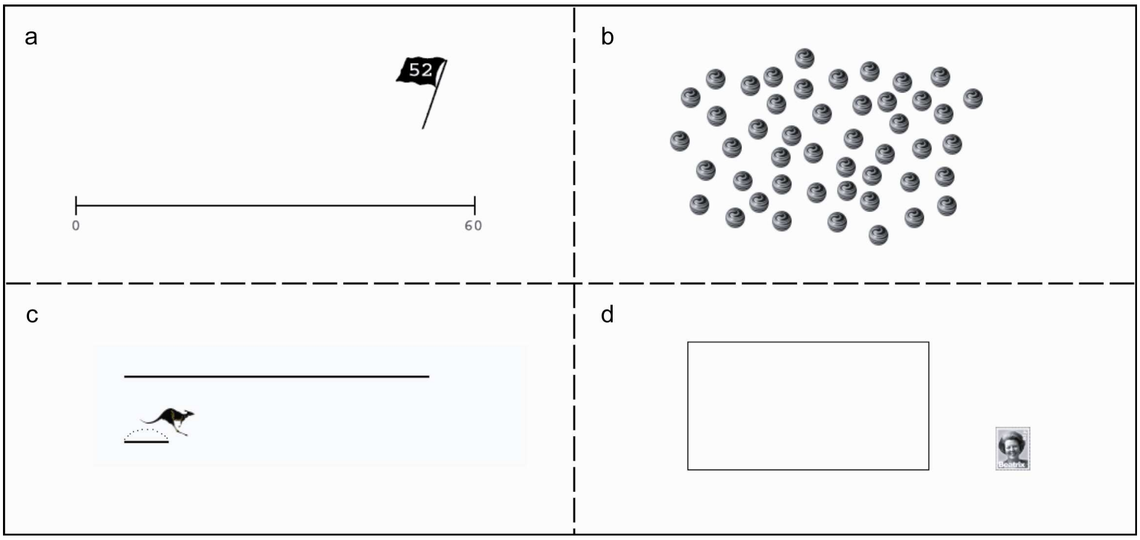

Figure 3.1. Examples of the problems in the a) number line, b) numerosity, c) length and d) area estimation task.

3.2.4 Procedure

The estimation test was administered to a whole class (of about 20 children) at the same time. The test was printed as an A5 booklet with one problem on each side of a page, i.e., two estimation problems were visible at the same time. Before each task, the experimenter explained the answering procedure to the children and provided two examples that were not included in the test. No clues or details about strategies were given. The children were told explicitly that they should only write down answers; it was not allowed to draw jumps (number line and length), or stamps (area), or to group marbles together as an aid to find the answer (numerosity). During the first three items the experimenter told the children when to turn to the next problem, in order to give an indication of the length of time in which they should finish a single problem (not more than 20 seconds). Thereafter, the children finished the task on their own. When most children were ready, the remaining children were asked to hurry up and if necessary the slowest children were allowed to skip a few problems. The data of most of these slow children were not further analyzed, see Analysis. The procedure for the numerosity task was slightly different from the other three tasks: after every 10-15 seconds the children were prompted to turn to the next problem in order to refrain them from counting. In spite of these precautions, the data showed that some of the children might have counted the marbles in the exercises with small numbers (11, 14, 16), because there were relatively many answers that were exactly correct. The three items were nonetheless preserved for analysis, because the reliability analysis (see Results) did not indicate that they should be treated differently from the rest of the items. The estimation test took between 20 minutes (in Grade 3) and 30 minutes (in Grade 1) to finish.

3.2.5 Mathematics education in Grade 1 to 3 in the Netherlands

Virtually all schools for primary education in the Netherlands use books for instruction that are based on Realistic Mathematics Education (RME). The fundamental idea of RME is that mathematics must be closely related to children’s everyday life experiences. Therefore, mathematical problems are often introduced within a “realistic” context. Furthermore, the process of learning mathematics is organized as “guided reinvention”: children are encouraged to discover mathematical principles through the use of their own informal strategies. There is a large emphasis on estimation in RME, but this usually involves finding approximate answers to arithmetic problems without doing an exact calculation (computational estimation) and not much time is devoted to other types of numerical estimation. Consequently, the children who participated in the present study presumably did not have experience with the estimation problems that were tested.

Although the children in our study came from schools that used different textbooks, they were all familiar with the use of the number line. From Grade 1 on, the “empty” number line is used as a tool for doing mathematics. Children learn to find the answer to additions and subtractions by drawing “jumps” on the empty number line; for example, 6 + 7 can be visualized with a jump of four units from the number 6 to the number 10, and then a jump of three units to the number 13. To learn the order of the counting sequence, children also practice the positioning of numbers on the 0-20 number line and later on the 0-100 number line. However, such a number line will usually show at least the decades, so there will be more marks than just the starting and the end point. Consequently, children had no experience with number line problems as tested in the current study. In the test, we used a 0-60 number line, because not all schools had used 100 as an end point yet in Grade 1.

Instruction in area and length measures starts at the end of Grade 2 or at the beginning of Grade 3. For example, children are taught how to measure length in centimeters and they receive instruction about the meaning of concepts like ‘kilometer’, ‘liter’, and ‘kilogram’. Measurement is treated only informally in Grade 3, so students are not yet expected to do calculations with area or length. There is no explicit training in estimating length and area.

Numerosity problems are regularly addressed in the instruction books, but the focus is usually on grouping objects together in order to facilitate counting, which is why drawing lines around the marbles was explicitly discouraged in the current study.

3.2.6 Analysis

The estimates of the numerosity, length and area problems were written down as numbers by the children themselves. The number line estimates had to be inferred from their marks on the 0-60 number line: the distance between the starting point of the number line and the child's mark was measured in cm (with a precision of 1 mm) and then multiplied with 5.

Initially, 273 children completed the estimation test. One child in Grade 1 was excluded from the analysis because her answers were clearly not related to the problems. Additionally, 16 children (5 in Grade 1, 5 in Grade 2, 6 in Grade 3) were rejected because they had too many missing items; the criterion for rejection was missing more than 3 items (25 %) on one of the four tasks. Missing items were mostly caused by children accidentally skipping a page, which was visible in the data as two consecutive missing items. In a few cases there was missing data because slow children were allowed to skip the last problems of a task, but after excluding the 17 subjects only 0.2 % of the estimates was located at the end of a task.

Before deciding on outliers, 8 estimates (of the 12,288) were removed because they were so far from the correct answer that including them would influence means and standard deviations immensely. For example, the answer '1000' when estimating 75 marbles. Outliers were defined separately for each grade: estimates that differed more than 2.5 standard deviations from the mean estimate of an item were removed. After removing outliers, another 19 children were rejected as subjects (6 in Grade 1, 9 in Grade 2, 4 in Grade 3) because they had too many missing items according to the criterion of missing at most 3 items in a task. The data set of the remaining 237 subjects had 2.5 % missing items in total (1.1 % missing, 1.4 % outliers).

Deviation scores relative to the magnitude of the correct answer were calculated: $|\,Estimate - CorrectAnswer\,| \;/\; CorrectAnswer$, and multiplied with 100 to obtain percent deviation scores. Then, for each estimation task the mean percent deviation on the 12 items was calculated. All analyses were conducted with the mean scores of the tasks, except for the reliability analysis. The reliability analysis was conducted with all 48 items; missing values were replaced with the child's mean score on the task. The alpha level was 5 % in all analyses of variance.

3.3 Results

To determine whether the estimation test provided a good measure of individual differences in estimation skills, a reliability analysis was carried out. The internal consistency for the 48 items was high; Cronbach's alpha was .84. Because there was only one item that would improve the alpha slightly if it was removed, it was decided to keep all items in the analyses. When the internal consistency was calculated separately for the different estimation tasks, Cronbach's alpha was .74 for number line, .80 for numerosity, .89 for length and .80 for area problems.

Table 3.1. Mean deviation scores, in percentages, for the three grades on number line, numerosity, length, and area estimation. The last row provides the general score on the estimation test.

	Grade 1		Grade 2		Grade 3	
	M	*SD*	*M*	*SD*	*M*	*SD*
Number line	48.5	23.7	23.6	9.5	22.5	9.2
Numerosity	32.3	12.3	23.3	9.9	20.1	9.0
Length	31.5	12.8	32.1	16.7	26.2	12.9
Area	42.9	13.1	41.1	15.8	32.8	13.5
Total estimation	*38.8*	*9.2*	*30.0*	*7.4*	*25.4*	*6.2*

Means and standard deviations for the percent deviation scores are provided in Table 3.1. From the Table, it is clear that estimation skills improved over grades. A repeated measures Anova with 'estimation task' as within-subjects factor, 'grade' as between-subjects factor, and 'percent deviation score' as dependent variable revealed a main effect of 'grade' ($F(2, 234) = 61.67$, $p < .001$, $\eta_p^2 = .35$), indicating that mean scores improved in higher grades. Post hoc Bonferroni tests showed that the general accuracy on the estimation test improved significantly from Grade 1 to Grade 2 (from 39 % deviation to 30 %, $p < .001$), and from Grade 2 to Grade 3 (from 30 % to 25 %, $p < .01$). Furthermore, a main effect of 'estimation task' was found ($F(3, 702) = 43.52$, $p < .001$, $\eta_p^2 = .16$), indicating that children scored differently on the four estimation tasks. Finally, there was a significant interaction effect between 'estimation task' and 'grade': $F(6, 702) = 16.99$, $p < .001$, $\eta_p^2 = .13$. Planned comparisons demonstrated a parallel development for length and area estimation ($p = .253$), whereas the development of all other tasks differed significantly from each other. In Figure 3.2, it is clearly visible that the development of the deviation scores was not the same for the different estimation tasks. Number line and numerosity scores improved significantly from G1 to G2 ($p < .001$ in both cases), but not from G2 to G3. In contrast, length and area estimation scores did not improve from G1 to G2, but did improve significantly from G2 to G3 ($p < .05$ for length and $p < .001$ for area estimation).

Although the current stimuli were not selected to make an optimal contrast, an attempt was made to find out whether the children had used a linear or a logarithmic representation of number magnitude. For each estimation task, the mean scores on the items were entered in a regression analysis and then analyzed separately for the three grades. The results showed that in all grades both linear and logarithmic representations fitted the data equally well.

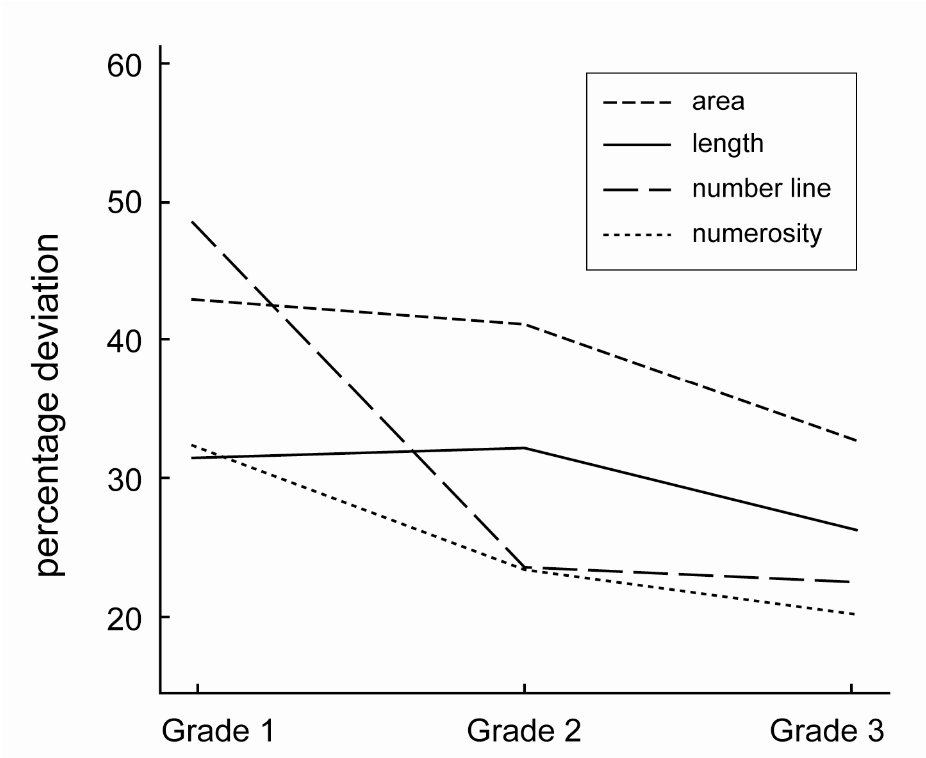

Figure 3.2. *Development over grades of the mean deviation scores (in percentages) on the estimation tasks.*

3.3.1 Principal components analysis

The relationships among the estimation tasks were explored by means of a principal components analysis, followed by a varimax rotation. The goal of the varimax rotation is to facilitate the interpretation of the calculated factors. When factor loadings are plotted in a scatter plot, the axes can be rotated in any direction without changing the relative locations of the points to each other. A rotation leads to a more simple factor structure, which makes it easier to determine the variables that explain most of the variance. In general, when factors are expected to be correlated, as is the case in the present study, a non-orthogonal rotation like oblimin is more appropriate than a varimax rotation. However, the calculated factors turned out to be hardly correlated (r = .17).

The principal components analysis revealed two components with eigenvalues greater than one, meaning that the components explained at least as much variance as one of the original variables would explain. Component I had an eigenvalue of 1.43 and explained 36 % of the variance. Component II had an eigenvalue of 1.21 and explained 30 % of the variance, for a cumulative variance explained of 66 %. The first component was well defined by the combination of number line and numerosity estimation, with component loadings of .83 and .85, respectively. The second component was well defined by the combination of length and area estimation, with component loadings of .80 and .75, respectively. See Table 3.2 for component loadings.

Table 3.2. Principal components analysis (with varimax rotation) of the subjects' mean deviation scores on the four estimation tasks.

	Component loadings	
	I	II
Number line	.83	.10
Numerosity	.85	–
Length	–	.80
Area	.13	.75

3.3.2 Correlations between estimation and mathematics

To investigate the relationships between estimation skill and general mathematical ability, the correlations between the two measures of mathematical ability and the deviation scores on the estimation tasks were calculated. The results, presented in Table 3.3, show that deviation scores for number line, numerosity and length estimation all correlated significantly with both CITO and TTR scores in Grade 1. However, in Grade 2 the correlations were limited to length estimation and number line estimation – the latter was only correlated with CITO. Finally, in Grade 3 no significant correlation coefficients between estimation scores and mathematical skills were obtained.

Table 3.3. Correlations between the scores on the estimation tasks and the two measures of mathematical ability.

	M	*SD*	CITO	Number line	Numerosity	Length	Area	Total estimation
Grade 1								
TTR	17.9	5.8	.61 **	-.41 **	-.31 **	-.28 *	ns	-.49**
CITO	41.5	15.0		-.45 **	-.35 **	-.24 *	ns	-.51**
Grade 2								
TTR	33.3	7.8	.64 **	ns	ns	-.37 **	ns	-.38**
CITO	65.7	8.6		-.23 *	ns	-.34 **	ns	-.37**
Grade 3								
TTR	39.9	8.3	.49 **	ns	ns	ns	ns	ns
CITO	79.6	11.4		ns	ns	ns	ns	ns

* $p < .05$, ** $p < .01$

3.3.3 Relative versus absolute deviation scores

In the present study, deviation scores relative to the magnitude of the correct answer were used. The choice for a relative measure was based on the desire to keep the tasks and the individual items as comparable as possible. This approach is different from Booth and Siegler's (2006); they calculated absolute deviation scores and corrected only for the scale of estimates (0-100 or 0-1000). We used relative deviation scores because we think this fits in best with what we know about numerical estimation in humans (Dehaene, 1997) and with how we deal with estimation in daily life: an error of 2 is judged more worrying on small quantities than on large quantities. Our data generally supported this view, because in three of the four estimation tasks answers were further from the correct answer on larger quantities, and also the variance in the scores was higher. An exception was number line estimation, perhaps because it was the only task with a known maximum. Especially in Grade 2 and Grade 3, children were more accurate on the quantities near the starting and the end point of the number line than on the quantities in the middle. This effect has been found before: in number line estimation, young children show the highest variance in their answers on the numbers near the midpoint of the number line (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008). Unfortunately, this implies that having used relative deviation scores lead to an overcorrection for large quantities on the number line task. To make sure that the present findings were not influenced by the choice of measure, all analyses were performed again with absolute deviation scores: $| Estimate - CorrectAnswer |$. Overall, this led to similar results, only this time in the repeated measures Anova it was found that number line and numerosity estimation had a parallel development, whereas the development of the other tasks differed from each other. From the other analyses the same conclusions can be drawn as from the analyses with relative scores.

3.4 Discussion

The main aim of the study was to gain more insight into the development of numerical estimation during the first three years of formal education. To this end, four different estimation tasks were tested in students of Grade 1, 2, and 3. We were interested in the development of the estimation scores over grades, but also in the relationships among the different types of estimation and the correlations between estimation scores and general mathematics scores. As a first step, a reliability analysis was carried out to evaluate whether the estimation test provided a good measure of individual differences in estimation skill. The results demonstrated a high internal consistency, not only for the separate tasks, but also for the estimation test in total. This suggests that the test provided a reliable measure of skill in estimation. Although the results of the reliability analysis confirmed that the scores on the different tasks were closely related to each other, a principal components analysis was carried out with the means per task to look into more subtle differences and similarities among the four estimation tasks. The results of this analysis indicated a difference between number line and numerosity estimation on the one hand, and length and area estimation on the other hand. Together, the two identified factors explained 66 % of the variance in the data.

The total score on the estimation test improved significantly over the three grades; deviation from the correct answer was 39 % in Grade 1, 30 % in Grade 2, and 25 % in Grade 3. When the tasks were studied separately, however, two different types of development were found. Number line and numerosity scores improved from Grade1 to Grade 2, but not from Grade 2 to Grade 3. In contrast, length and area scores showed no improvement from Grade 1 to Grade 2, but did improve from Grade 2 to Grade 3. On the whole, children performed better on length than on area problems. This is not surprising because of the extra spatial dimension that is involved in area estimation. Nevertheless, the course of the development was similar in the two domains. A repeated measures analysis confirmed the parallel development for length and area estimation. The results showed no parallel development for number line and numerosity estimation, but this may be due to a scaling problem; when absolute deviation scores were used instead of relative scores, a parallel development was found for number line and numerosity estimation, but not for length and area.

3.4.1 Relationships among estimation skills

In the present study, the results of the PCA showed that skill in estimation for number line and numerosity problems can be distinguished from the skills needed for length and area estimation. This was also reflected in the development of the estimation scores over the grades: number line and numerosity estimation followed a different developmental course than length and area estimation. What could explain these different developmental trajectories? A plausible explanation for the similarity in performance between length and area estimation is that the most appropriate strategy to solve the problems is a benchmark strategy. A child might try to fit the reference unit visually into the larger unit: how many small units are needed to cover the larger unit? In the case of number line and numerosity estimation something else is needed, namely a good understanding of the quantity of a number.

The largest improvement in performance was between Grade 1 and Grade 2 on the number line scores. Grade 1 children performed very poorly on number line estimation; their scores deviated nearly 50 % from the correct answer and the variance in scores was very high. In contrast, children in Grade 2 scored only 24 % deviation from the correct answer. How do we explain this large improvement? It is known from other studies that young children prefer a counting strategy to obtain number line estimates (see Siegler & Booth, 2005). Often, they count upwards from zero with an arbitrarily chosen unit for ‘1’. This counting strategy is not only inefficient, it is also very time consuming. In the present study, children were discouraged from using a counting strategy: the experimenter explained there was only a short time to think and that they should pick the location that they thought was ‘more or less’ right. However, Grade 1 children may still have relied largely on counting strategies. Petitto (1990) showed that children in Grades 1 to 3 move progressively from number line strategies that are strictly sequential to strategies that incorporate elements of proportional reasoning. This explains the large improvement in the scores from Grade 1 to Grade 2. Children in Grade 2 will have learned to make use of anchor points like the midpoint or the quarter values, especially in the Netherlands where the number line is a commonly used tool in mathematics instruction.

The differences in development over grades suggest that skills needed for number line and numerosity estimation develop earlier than skills necessary for length and area estimation. In fact, there was no significant improvement for number line and numerosity scores between Grade 2 and Grade 3. Does this mean that children have reached the highest performance for number line and numerosity estimation already in Grade 2? This seems unlikely: children in Grade 2 and Grade 3 still had a deviation score of over 20 % for number line and numerosity, which leaves much room for improvement. Previous research has shown that performance on number line and numerosity estimation improves steadily throughout elementary school, although the development is slow and gradual. Siegler and Booth (2005) list several sources of difficulty that could explain the slow development of numerical estimation: limitations of conceptual understanding, limitations of skills such as counting and arithmetic, and limitations of working memory. Perhaps then improvement is still possible after Grade 2, but at a slow rate. Of course, this would have to be verified by means of a longitudinal study that takes also higher grades into account.

3.4.2 Correlations between estimation and mathematics

The correlations between the mean score per task and the individual mathematics scores revealed a positive relationship in Grade 1 between general ability in mathematics and skill in number line, numerosity and length estimation. However, only length and number line estimation were correlated with mathematical ability in Grade 2, and in Grade 3 no correlations were present at all. This was an unexpected finding, given previous research (Booth & Siegler, 2006). Part of the reason for the disappearance of correlations in higher grades could be that the variance in the estimation scores decreased: the standard deviation of the total estimation test was 9.2, 7.4, and 6.2, respectively – although the decrease in variance between Grade 1 and Grade 2 was mainly caused by the number line task. Therefore, the estimation test might not be optimally designed to detect relations between estimation skills and mathematical ability in older children. However, if the disappearance of the correlations in Grade 2 and Grade 3 reflects a genuine effect, it shows that mathematical skill and estimation skill develop in unison in Grade 1 and that this relationship becomes weaker in higher grades.

Of course, it is difficult to draw conclusions on causality from correlations between estimation skill and mathematical ability in Grade 1 and Grade 2. First of all, it is possible that the correlations were caused by a common source, like processing speed or general intelligence. Future studies could include a factor like ‘processing speed’ to evaluate the possibility that correlations are caused by a general underlying skill instead of skills specifically needed for mathematics or estimation. Second, if there is indeed a causal relationship between skill in mathematics and skill in estimation, we do not know the direction of the effect, because both directions are equally plausible. Good conceptual knowledge of numerical magnitude might be beneficial for learning mathematics, especially when children are still struggling with the fundamental principles of adding and subtracting. On the other hand, children with good mathematical skills will have an advantage in estimation problems, for example when they try to find the midpoint or the quarter values of a number line. Most likely, experimental training studies are required to dissolve the issue on causality in the relationships between mathematical ability and estimation skills.

3.4.3 Methodological limitations

There are some factors that were not appropriately controlled for in the present study. Most important, the range of the number sizes was not the same across the four estimation tasks. The fact that the answers of the number line and numerosity problems were larger in size than the answers of the length and area problems may have influenced the classification of the different estimation types. Therefore, we cannot rule out the possibility that the observed difference between length and area estimation on the one hand, and number line and numerosity estimation on the other hand, is actually an artifact. The estimation tasks differed from each other also on other aspects: the numerosity task had a more strict time limit than the other three tasks, and the number line task required an estimation of relative magnitude, whereas the other tasks required estimations of absolute magnitude.

There were also some general problems with the experimental procedure. First of all, the order of the tasks and the order of the problems within a task was not counterbalanced. Second, it would have been better if only one problem instead of two problems was visible at the same time, because children might have inferred their estimates on the right page from their estimates on the left page. Third, we did not check whether the youngest children were able to write down the values of large numbers. Young children often have place value problems; for example, they write down “36” when they mean “63”. Problems with writing down numbers, which was required in three of the four estimation tasks, may have influenced age related findings. Finally, the choice for the less regular end point of “60” in the number line task may also have influenced the results and the age related findings for this task.

3.4.4 Conclusions

The current results suggest that using a principal components analysis is a promising approach to detect the underlying structure of human performance on different tasks, because the two identified factors explained a large part of the variance in the scores of the children. Therefore, PCA might be applied more often in studies on children's behaviour when a range of different tasks is involved. This technique could help to find associations between variables that otherwise might go unnoticed.

Overall, no direct educational implications can be inferred from the results of the present study because we do not know, for instance, whether training children in numerical estimation will help them with their performance on mathematical problems. However, we observed that children could use some help in learning numerical estimation, because even in Grade 3 children scored an overall deviation of 25 % from the correct answer, whereas they should be comfortable with the range of numbers tested in the estimation. Estimation is an important skill in daily life, so there seems to be a need for children to improve their estimation skills.

In summary, the results of this study show that estimation skill is to a large extent a general skill that can be measured quite reliably by means of different tasks. However, when looking closely at the relations among the tasks and the development of the scores over the grades, differences were found between number line and numerosity estimation on the one hand, and length and area estimation on the other hand. This suggests that different procedural skills are needed for the two groups of estimation tasks. Furthermore, it can be concluded that estimation skill and general mathematical ability are more intertwined in younger children than in older children. There are also some limitations of the experimental design used in the present study, so the conclusions are only tentative. Further research is needed to study the development of estimation skills and the relation between estimation and mathematics.

3.5 Acknowledgments

We are grateful to all the students and teachers who participated in this study.

4

Learning basic addition facts from choosing between alternative answers

The acquisition of addition facts was investigated in a practice study. Participants were 103 Grade 1 children who practiced simple addition problems with three different methods: (a) writing down the answer, (b) choosing between two alternative answers, and (c) filling in the second missing addend. On a test with simple addition problems, children who practiced with the Choice method showed positive transfer: choosing between two answers was about as effective for learning addition facts as the conventional method of writing down the answer. There was no transfer effect for children who practiced with the Missing Addend method. The results are in accordance with network theories on arithmetic fact learning and specifically the Identical Elements (IE) model of arithmetic fact representation.

Published as: Van Galen, M. S., & Reitsma, P. (in press). Learning basic addition facts from choosing between alternative answers. *Learning and Instruction*.

4.1 Introduction

Over the last decades, there has been much discussion about the best way to teach children mathematics: should they learn skills first or concepts first? Proponents of the skills-first approach (Briars & Siegler, 1984) argue that children need to learn *how* to do mathematics first, and through applying their skills they will discover arithmetic regularities and concepts. In this approach, arithmetic procedures are taught in an explicit way and there is a focus on practice and reinforcement. In contrast, proponents of the concepts-first approach (Carpenter, 1986) argue that children should learn mathematics in a meaningful way and that it is important that children realize *why* a certain arithmetic procedure can be used to solve a mathematical problem. In this approach, extensive practice is postponed until children have a good conceptual understanding of mathematics. Both sides agree, however, that at some point during education children need to practice mathematical procedures in order to gain routine expertise. This is especially the case with basic arithmetic problems. Learning to solve basic arithmetic problems accurately and with little effort is an important goal in primary education, because fast retrieval of answers to simple problems (e.g., 3 + 6 = …) is a prerequisite for solving more complex problems (e.g., 13 + 26 = …) (Cumming & Elkins, 1999).

In general, adults solve addition facts with sums up to 10 using direct retrieval from long-term memory, but even for these simple problems backup strategies like counting and transformation to known facts (e.g., 3 + 4 = 3 + 3 + 1) have been reported (Campbell & Austin, 2002; LeFevre, Sadesky, & Bisanz, 1996). Through practice with addition problems, children in primary school move on from initial counting procedures to faster memory-based strategies. Frequent repetition of problems is an important factor in this process (Ashcraft & Christy, 1995; Geary, 1996). Surprisingly, it is unknown which type of exercises is most effective to obtain automaticity in addition facts. In the present study, we introduce a new method of practicing addition facts: choosing from two alternative answers. We will evaluate the effectiveness of this method and compare it to the conventional way of answering addition problems. We will also evaluate another method of practicing: filling in the missing second addend of an addition problem. We will explore how much children learn from this method and whether there is a transfer of knowledge to regular addition problems. Implications for educational practice will be discussed.

4.1.1 Learning basic arithmetic facts

In adults, knowledge of arithmetic facts is assumed to be organized as an associative network in long-term memory (Ashcraft, 1992; Campbell, 1995; McCloskey, Harley, & Sokol, 1991; Siegler, 1988; Verguts & Fias, 2005). In such a network various connections exist between a problem and possible solutions. When a problem is solved correctly on repeated occasions, the association between the problem and the correct answer becomes stronger, that is, the probability for correct retrieval increases. There is much empirical evidence for storage of arithmetic facts in an associative network, coming from studies on production errors (Campbell, 1994) and on interference and priming effects (Campbell, 1987; Campbell & Clark, 1989). Siegler (1988) assumed that children initially solve multiplication problems by nonretrieval strategies, such as repeated addition. The answers found with nonretrieval strategies (both correct and incorrect) shape the later associative network. If sufficient associative strength has accumulated between a problem and one or more answers, the answer is retrieved directly from memory. This is not to say that direct retrieval always leads to the correct answer, because there may be interference from competing false solutions. Multiplication errors are, for instance, often table related (e.g., $3 \times 5 = 18$ instead of 15). However, with practice and feedback, “peaks” will develop in the distributions of associations such that the strongest associations are between arithmetic problems and their correct answer.

In a study on automatic activation of arithmetic facts (Lemaire, Barrett, Fayol, & Abdi, 1994; Experiment 1 and 2), evidence was found that children of elementary school activate addition facts automatically, just like adults. When asked to indicate whether a probe had been present in a previously viewed number pair, children from different ages took more time to reject distractors equal to the sum of the pair than unrelated numbers. However, the interference effect depended on the size of the numbers in the pair and the age of the child. Second-graders (7.1 years of age) showed an interference effect only when both integers in the original pair were 5 or smaller. Third-graders (8.2 years of age) also showed an interference effect when one integer in the original pair was 5 or smaller and the other was between 6 and 9. And like adults, fourth-graders (9.0 years of age) and fifth-graders (10.4 years of age) showed an interference effect on all presented problems. These findings are consistent with the view that children gradually develop an associative network for arithmetic facts throughout primary school.

In a Dutch study (Ruijssenaars, Van Vliet, & Willemse, 2002) it was found that for learning multiplication facts it is sufficient that children recognize on several occasions that a problem and a solution are related. A group of children practiced multiplication facts with a method that involved choosing the correct answer to a problem out of two possible answers (e.g., $6 \times 7 = 45$ or 42). Although this seems to be a quite “passive” way of gaining knowledge on arithmetic facts, children who practiced with this method actually showed a large improvement on multiplication problems presented in the traditional way ($6 \times 7 = \ldots$). In fact, their performance was almost as good as that of children who had practiced the problems in the same way as they were tested. Apparently, choosing the correct answer strengthens the connection between an arithmetic problem and its solution effectively. The present question then is whether this method of practicing is also effective for learning addition facts.

In the last decades, research on arithmetic facts has established that mechanisms involved in storing addition and multiplication facts in memory are highly similar (for an overview on performance characteristics in simple addition and multiplication see Campbell, 1995). For instance, both addition and multiplication problems suffer from the problem-size effect: people take longer and make more errors if the digits in an arithmetic problem (operands and answer) are larger in numerical size. Also, tie problems with a repeated operand are easier to remember than non-tie problems, both in multiplication (e.g., 3×3) and in addition (e.g., $4 + 4$). Although there is evidence that addition and multiplication facts are stored in separate semantic memory networks (Van Harskamp & Cipolotti, 2001), the finding that some production errors stem from choosing the wrong operation (e.g., $6 \times 7 = 13$; Miller, Perlmutter, & Keating, 1984) suggests that these networks are interrelated. Overall, because mechanisms guiding the learning of arithmetic facts appear highly similar for addition and multiplication, we expect that choosing between alternative answers could also be a fruitful approach for learning basic addition problems.

4.1.2 Verification tasks

The method of choosing between two alternative answers as used in the present study is actually very similar to verification tasks regularly used in studies on mental arithmetic. In these tasks participants make decisions on whether the proposed answer for a stated problem is true or false ($8 + 4 = 13$?). Although it is often assumed that verification tasks are performed by producing an answer and comparing it with the presented answer, Zbrodoff and Logan (1990) proposed that verification involves comparing the equation as a whole against memory. Participants may not always compute or retrieve the exact answer, but they make a rough judgment of whether the presented combination of problem and answer looks familiar. Additional evidence for this hypothesis comes from a brain-damaged patient whose accuracy on arithmetic facts was far better for subtraction than for addition and multiplication, at least when measured in a production task (Dagenbach & McCloskey, 1992). In a verification task, however, accuracy was similar across different operations. These results suggest that although the patient was generally unable to recall the correct answers to addition or multiplication problems in the production task, these problems evoked some feeling of familiarity when presented in the verification task. Verification tasks may therefore reveal tacit knowledge. The question is of course whether a verification task could also enhance knowledge about number facts when this method is used as a method of practice during a certain period of time.

An advantage of using a verification task as a method of practice is that children presumably solve verification problems faster than problems presented in a conventional format. If addition problems are practiced within a fixed period of time, choosing between alternative answers could even be more effective than writing down the answer to a problem, because children will probably solve more problems. It has been shown that the frequency with which arithmetic facts occur in elementary school books influences the time needed for retrieving the answer (Ashcraft & Christy, 1995; Geary, 1996). Furthermore, network theories of arithmetic fact retrieval predict that finding the correct solution on many occasions increases the probability that the correct answer is retrieved the next time a problem is encountered. Therefore, if children solve more problems with the method of choosing between answers, it is possible that this method is more effective than writing down the answers to problems.

However, it is also possible that choosing between alternative answers is less effective than answering problems in the conventional way because children do not have to compute exact answers, they may merely guess. The attention they devote to choosing the correct answer may be cursory and therefore less effective for the storage of addition facts in memory. Furthermore, by presenting two alternative answers, children are not only confronted with the correct answer of the problem, but also with an incorrect answer. If arithmetic facts are stored as an associative network in memory, presenting children with an incorrect answer could strengthen the association between the problem and the incorrect answer. This could be especially troubling if the problems are relatively novel to the children. Therefore, there are also reasons to expect less positive effects of the method of choosing between alternative answers.

4.1.3 Transfer effects from practice

In the study of Ruijssenaars et al. (2002) a transfer of knowledge was found from one presentation format to another, namely practicing problems with two alternative answers led to an improved performance on regular multiplication problems. However, Ruijssenaars et al. (2002) also studied another method of practicing multiplication facts. This method involved filling in a missing factor, for example, 6 × … = 42. Practicing with this method did not improve children’s performance on problems that were presented in a regular format (6 × 7 = …). In other words, there was no transfer effect. This finding is in accordance with a study in adults by Rickard and Bourne (1996), who also found that practicing problems with a missing factor did not lead to a better performance on regular multiplication problems. Based on their findings, Rickard and Bourne developed the Identical Element (IE) model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996; Rickard, Healy, & Bourne, 1994). According to the IE model, there is a single long-term memory node for problems consisting of the same numerical elements (i.e., operands and answer), regardless of operand order. This implies that when the numerical elements of a test problem do not match exactly with those of a practice problem, there will be no positive transfer. Rickard and Bourne (1996) found positive transfer when multiplication problems were presented with an operand order change (e.g., 6 × 7 = … practiced, 7 × 6 = … on the test), but not when the numerical elements were different (e.g., 6 × … = 42 practiced, 6 × 7 = … on the test).

In the present study, a similar method of practice involving filling in a missing addend (e.g., 3 + … = 5) was used. Foremost, this is a control condition for the method of choosing between alternative answers, because a positive transfer effect to regular addition problems was expected when children practice with the Choice method and *not* when children practice with the missing-addend method.

However, a transfer effect to control problems cannot firmly be precluded. Campbell, Fuchs-Lacelle, and Phenix (2006) showed in a priming study that the identical element (IE) model applies not only to multiplication, but also to addition and subtraction. However, they also found that the IE model does not tell the whole story on transfer effects in adults. For instance, there was a small, but nevertheless significant transfer effect for subtraction problems with an operand order change (e.g., 15 – 9 = … was presented first, 15 – 6 = … was presented a few trials later). Such effects can only be reconciled with the IE model if procedural routes are also taken into account. Therefore, "associative transfer" (resulting from strengthening a common problem node) should be distinguished from "mediated transfer" (resulting from strengthening a related, mediator problem). For instance, the problem 15 – 9 could strengthen the mediator problem 9 + 6, leading to reaction time (RT) savings on the problem 15 – 6. Campbell et al. (2006) did not include missing-addend problems, so we do not know whether adults show positive transfer from missing-addend problems to regular addition problems. However, the route of mediated transfer leaves open the possibility of a transfer effect.

4.1.4 The present study

In the present study the effectiveness of two methods of practice for learning basic addition facts was evaluated: the Choice method, in which the problems used required choosing between two alternative answers (e.g., 3 + 2 = 5 or 7), and the Missing Addend method, in which the problems used required filling in the second missing addend (e.g., 3 + … = 5). The two methods were compared to what we call the Answer method, in which the problems used required writing down the answer (e.g., 3 + 2 = …); this is the traditional way of practicing addition problems.

4.1.5 Hypotheses

With respect to the results during the practice period, it was predicted that the Choice method will be easier than the Answer method, whereas the Missing Addend method will be more difficult than the Answer method. Therefore, during practice children in the Choice condition were expected to be faster (Hypothesis 1a) and more accurate (Hypothesis 1b) than children in the Answer condition, whereas children in the Missing Addend condition were expected to be slower (Hypothesis 2a) and less accurate (Hypothesis 2b) than children in the Answer condition.

With respect to the results of the tests after practice, for children in the Choice condition a learning effect (a significant difference between scores for practiced and unpracticed problems) was expected on a test with problems presented in the Choice format (Hypothesis 3a), as well as a transfer effect on a test with problems presented in the Answer format (Hypothesis 3b). For children in the Missing Addend condition the prediction was that there will be a learning effect on a test with problems presented in the Missing Addend format (Hypothesis 4a), but no (or only a marginal) transfer effect on a test with problems presented in the Answer format (Hypothesis 4b).

Finally, it was also explored whether learning effects were different for children with poor, average, and good mathematical ability. The Choice method was expected to be especially beneficial for students with poor mathematical ability, because practicing with this method is easier than practicing with answer problems. In the practice sessions, children solve as many problems as they can within a particular time period, which means that children using the Choice method will probably solve more problems than children using the Answer method in that period. This may be especially true for children with poor mathematical ability, because it is possible that they have difficulties in finding the answer to a problem when presented in the Answer format, but recognize the correct answer when presented in the Choice format. Because frequent repetition is important to gain speed in solving basic arithmetic problems, the following result for the children with poor mathematical ability was expected: on a test with problems presented in the Answer format, the transfer effect from children who practiced with the Choice method will be larger than the learning effect from children who practiced with the Answer method (Hypothesis 5a). Furthermore, it was expected that if a transfer effect occurs for children in the Missing Addend condition, this effect will be found only in children with good mathematical ability (Hypothesis 5b).

4.2 Method

4.2.1 Participants – design

Participants were 103 Grade 1 students (48 girls and 55 boys) from five Dutch schools for primary education. Mean age of the children was 6.9 years (*SD* = .42). For the purpose of analyzing the data of the study, the children were divided into three groups according to their score on a standard mathematics test (Central Institute for Test Development [CITO], 2005). This test is part of a student monitoring system in the Netherlands and was administered according to an established schedule in the schools a few weeks after the pretest. Children who scored under the 25th percentile of the normalized scores were classified as ‘poor’ (28 students), children between the 25th and 75th percentile were classified as ‘average’ (52 students), and children above the 75th percentile were classified as ‘good’ (23 students).

Participating children formed three practice groups: the Answer group, the Choice group, and the Missing Addend group. Each group practiced for three weeks on a small set of problems with the respective method of the three different training methods. After the practice period, participants were tested on practiced and unpracticed items, presented in all three formats. This design enabled us to study two different types of learning effects. First, we could establish whether children actually learned from practice – this would be the case if they scored higher on practiced compared to unpracticed problems on a test with problems in the same format as the problems they practiced. Second, we were able to study transfer effects. If a learning effect occurs on a test with problems presented in a different format, this means there was a transfer of knowledge to problems in the unpracticed presentation format.

Almost all schools for primary education in the Netherlands use books for instruction that are based on Realistic Mathematics Education (RME). The basic idea of RME is that learning mathematics is a constructive activity and should be closely related to children's everyday life experiences. Children are encouraged to use different strategies to solve a problem and the teacher can use the strategies of the children as a starting point for a classroom discussion. Blöte, Klein, and Beishuizen (2000) describe an experimental RME program in which as much as one-third of the mathematics lesson is spent on discussion. However, in a normal classroom situation the amount of time spent on discussion is much lower, because the regular textbooks for mathematics instruction do not expect teachers to incorporate that much discussion in their mathematics lessons. In general, RME offers what Baroody (2003) calls an investigative approach to teaching mathematics, with a focus on meaningful acquisition of skills and the development of mathematical thinking, under the guidance of the teacher.

Although the children in the present study came from schools that used different textbooks, the mathematics curriculum is roughly the same on every school. At the moment of the pretest participants had received about five months of schooling in Grade 1. They were familiar with the format used in the Answer condition, but not with the other two formats. Dutch textbooks never present problems in the format of the Choice condition. The format of the Missing Addend condition is sometimes used in textbooks, but only in higher grades. At the period of testing participants were familiar with the difficulty level of the problems used in this study (addition up to 10) and could solve the problems on their own, but the problems had not been practiced intensively in class, that is, no special arrangements were provided for repeatedly practicing the addition problems.

Originally, 116 children participated. Five children were excluded from the data set (three from the Answer condition and two from the Missing Addend condition) because they were absent on more than two of the nine practice sessions. Additionally, eight children in the Missing Addend condition (six with poor and two with average mathematical ability) stopped practicing prematurely because the problems were too difficult for them, according to their teachers. As a consequence, fewer children practiced in the Missing Addend condition than in the other conditions, that is, 35 children practiced in the Answer condition, 39 in the Choice condition, and 29 in the Missing Addend condition.

Because most children who stopped practicing in the Missing Addend group had poor mathematical ability, we expected the average level of mathematical ability to be higher in the Missing Addend group than in the other two groups. However, the composition of the Missing Addend group (24% poor, 55%

average, 21% good) did not appear to be unbalanced, compared to the Answer group (28.5% poor, 51.5% average, 20% good) and the Choice group (28% poor, 46% average, 26% good). Statistically, there were no significant differences between the groups in their composition, $\chi^2(4, N = 103) = .73, p = .95$.

Furthermore, we compared the three practice groups on their mean score on the pretest. No differences were found between the groups, $F(2, 100) = .07, p = .93$. The practice groups were also compared on reading level. Reading scores came from a standard reading test (CITO, 1992) and were obtained from the schools. For two children the score on the reading test was missing in the school's registration system (one from the Answer condition and one from the Choice condition), so we compared the scores of 101 children. No differences were found between the three practice groups, $F(2, 98) = .45, p = .64$.

4.2.2 Material and procedure

Two different training sets (Set A and Set B) were constructed, each consisting of 13 addition problems with outcomes ranging from 5 to 10. The numbers used as first addends in Set A were also used as first addends in Set B, and likewise for second addends and outcome numbers. Problems and fillers are listed in Appendix A. Participants were tested three times on their proficiency on the two sets: (a) pretest (before the practice period), (b) posttest (one to three days after the last practice session), and (c) retention test (one month after the practice period).

Two paper and pencil two-minute tests were administered to a whole class at the same time (12-22 children). The testing started with the problems of Set A. The participants were instructed to write down the answers to as many problems as possible within 2 minutes. The 13 problems of Set A were repeated five times on the two-minute test (in different orders) to make sure that the participants would not be able to finish the problems within the time limit. After Set A, Set B was tested in the same way.

In each school, the children were divided into three practice groups that were about equal in mathematical ability according to their score on the pretest (number of problems solved correctly)[1]. Each group practiced in a different training condition: (a) Answer condition (participants wrote down the answers to the addition problems), (b) Choice condition (participants chose one of the two alternative answers by drawing a circle around the correct answer), and (c) Missing Addend condition (the answer of the problem was already given, but participants had to complete the problem by writing down the second addend). In each condition half of the children practiced on Set A and the other half practiced on Set B. (See Figure 4.1 for an example problem in each condition.)

---

[1] The results of the standard mathematics test were not yet available then, therefore we used the score on the pretest to create three practice groups that were about equal in mathematical ability.

3 + 2 = ______
3 + 2 = 10 5
3 + ______ = 5

Figure 4.1. Examples of the problems in the three training conditions. From top to bottom: Answer condition, Choice condition, Missing Addend condition.

Training material was A4 size booklets with 1350 problems, built up from 90 blocks of 15 problems. Each block consisted of 13 experimental problems and two filler problems. The order of the problems within a block was pseudo-random, namely in consecutive items the same number was used at most two times on the same position in the problem (first addend, second addend, and outcome). In addition, the first problem of a block was not the same as the last problem from the previous block, and fillers were not listed directly after each other. The order of the problems was the same in every condition. Problems were presented in groups of five, with spaces in between. The booklets of the Answer and the Missing Addend condition had 50 problems on each page, in two columns. The booklets of the Choice condition had 45 problems on each page, in three columns. In the Choice condition, after the "=" two alternative answers were printed above each other, see Appendix A for the incorrect answers. The correct answer was the upper alternative in half of the cases and the lower alternative in the other half. In consecutive problems, the position of the correct alternative was at most three times up or three times down.

There were several constraints for the selection of the incorrect answers in the Choice condition. The alternative answer had to be reasonably acceptable for the children and, therefore, the alternative number was larger than the first addend and was not the same number as the second addend. Just like the correct answers, the incorrect answers were not larger than 10. Each problem had three incorrect answers – with the exception of 7 + 2 and 7 + 3, because it was not possible to select three incorrect answers within the constraints specified before. The incorrect alternatives consisted of both even and odd numbers. Each of the three incorrect answers was randomly assigned to the problem in one-third of the cases. Filler problems – that appeared less often in the material – had two incorrect answers instead of three. Over the whole set of incorrect alternatives, the variation in the numbers used was as large as possible.

In the posttest, participants made four additional two-minute tests. Just like the pretest (before the practice period) and the retention test (one month after the practice period), they started with the problems of Set A and Set B printed in the format of the Answer condition. Subsequently, the problems of Set A and Set B were also tested in the other two formats: first in the format of the Missing Addend condition and then in the format of the Choice condition. Note that the choice problems and missing-addend problems were only administered at the posttest. Table 4.1 gives an overview of the two-minute tests made on each test occasion.

Table 4.1. Overview of the two-minute tests at the pretest, posttest, and retention test.

Pretest	Posttest	Retention test
1. Answer (Set A)	1. Answer (Set A)	1. Answer (Set A)
2. Answer (Set B)	2. Answer (Set B)	2. Answer (Set B)
	3. Missing Addend (Set A)	
	4. Missing Addend (Set B)	
	5. Choice (Set A)	
	6. Choice (Set B)	

4.2.3 Training procedure

All children in a school class participated. Each child received a personal booklet for training. For three weeks they practiced three times a week with their own teacher (Sessions 1-9); each practice session lasted for 10 minutes. At the end of the week the booklets were corrected by the experimenter, who made a mark if children made a mistake or forgot to fill in the answer. In Sessions 4 and 7 the children first studied their mistakes and tried to correct them before they carried on with solving new problems for the remaining time of the 10 minutes. After three weeks of practice there was one final session (Session 10) in which the children only corrected their mistakes and did not practice. For the Choice condition correcting a problem simply meant marking the other alternative.

The teacher instructed the children how to do the exercises: the problems had to be solved from top to bottom, before turning to the next column. The children were to solve all of the problems, without missing out any. Some of the children used additional materials to solve the problems, for instance an abacus or the number line with the numbers from 1 to 10. This was allowed during the practice period, because we wanted participants to solve the problems as accurately as possible. After 10 minutes of practice, the teacher told the children to draw a line under the last problem they solved and to write down a number which the teacher provided and that corresponded to the practice session. In the next session the children started right where they ended in the previous session. There was no limit on the number of problems that participants could make in one session. Six children did not start any new problems in Session 4 or in Session 7 because they needed all of the 10 minutes to correct their errors of previous sessions. All children started with new problems in Sessions 5 and 8, irrespective of whether they had finished correcting in the previous session.

4.2.4 Analysis

Both the practice and the test data were analyzed. The practice data is based on 99 participants instead of 103, because four children showed practice results that were different from the other children and including them could obscure general patterns. Three of these four children (two in the Answer condition and one in the Choice condition) managed to answer all 1350 problems in the booklet before the end of the practice period; their data is not included because the number of problems made was below their capacity in the third week. Furthermore, one child in the Missing Addend condition mistakenly answered the problems in the first two practice sessions by adding the first addend to the outcome number. We did not exclude this child from the experiment altogether because in the remaining sessions he answered the problems properly. Because these four children practiced sufficiently to be included on the results of the tests, the analysis on the pretest, posttest, and retention test will include all 103 participants.

4.3 Results

4.3.1 Number of items answered during practice

Table 4.2 summarizes how many problems the children answered on average in each week of practice, presented separately for children with poor, average, and good mathematical ability. The filler items are not included in these figures, so the children actually solved more problems. It should be noted that the time available for answering problems was reduced in the second and third week compared to the first week, because children started the first session of the week with correcting their mistakes. Note also the high standard deviations in Table 4.2, showing that individual differences between children were large.

Table 4.2. Means and standard deviations of the number of items answered during practice for the three training conditions as a function of week of practice and children's mathematical ability.

	Week 1		Week 2		Week 3		Total	
Level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Answer condition								
Poor	100	53	95	60	97	56	292	145
Average	129	55	116	53	119	54	365	145
Good	182	73	153	73	178	75	513	206
Total	130	62	116	61	123	63	369	170
Choice condition								
Poor	125	58	119	55	129	65	374	153
Average	184	95	173	113	126	68	482	244
Good	265	144	231	115	220	104	716	345
Total	186	110	171	106	149	85	506	275
Missing Addend condition								
Poor	57	33	68	36	54	27	179	85
Average	105	58	120	81	128	80	352	205
Good	127	87	138	57	162	59	427	189
Total	97	64	111	70	117	75	325	196

Filler items are not included.

A repeated measures ANOVA was carried out with Week (1, 2, 3) as within subjects variable and Training (Answer, Choice, Missing Addend condition) and Level (poor, average, good mathematical ability) as between subjects variables. The dependent variable was the number of problems answered in each week of practice. There was a main effect of Level, $F(2, 90) = 10.17$, $p < .001$, partial $\eta^2 = .18$, namely children with good mathematical ability had higher scores than children with average ($p < .01$) and poor mathematical ability ($p < .001$). There was also a main effect of Training, $F(2, 90) = 7.73$, $p < .01$, partial $\eta^2 = .15$. Post hoc Bonferroni tests showed that children in the Choice condition answered significantly more items than children in the Answer condition ($p < .05$) and children in the Missing Addend condition ($p < .01$). There was no significant difference between children in the Missing Addend and the Answer condition ($p = 1.00$). Finally, there was an interaction effect between Week and Training, $F(4, 180) = 3.43$, $p < .05$, partial $\eta^2 = .07$ (with a Greenhouse-Geisser correction), showing that the pattern of answered items over the weeks was different for the three training conditions. Looking at the total scores for the training conditions in Table 4.2, we see that in the Answer condition the number of answered items in the practice weeks developed as expected. Specifically, the number of answered items in Week 2 was lower than in Week 1, because less time was available to solve new problems, but in Week 3 the number was higher than in Week 2, because children became faster in answering the problems. For children in the other two training conditions the development during practice was different: children in the Choice condition showed a decrease over the practice weeks, whereas children in the Missing Addend condition showed an increase.

By applying within subjects contrasts it was tested whether there was a significant difference in each of the three training conditions between Week 1 and Week 2 and between Week 2 and Week 3. The only significant difference was between Week 1 and Week 2 in the Missing Addend condition, $F(1, 27) = 4.24$, $p < .05$, partial $\eta^2 = .14$. This increase was probably due to participants needing time to understand the format; teachers shared their observations with us that the children initially considered the missing-addend problems quite difficult. The other within subjects contrasts were not significant. However, over the whole practice period there was a significant difference between the children in the Choice condition (who answered 37 problems less in Week 3 compared to Week 1) and children in the Missing Addend condition (who answered 20 problems more in Week 3 compared to Week 1); posthoc Bonferroni tests showed a significant difference between the two groups on the difference between Week 1 and Week 3 ($p < .01$). There was no significant difference between the Answer condition and the other two training conditions.

4.3.2 Percentage of errors during practice

Table 4.3 shows the percentage of errors in each week of practice (number of errors / number of answered problems × 100), presented separately for children with poor, average, and good mathematical ability. A repeated measures ANOVA was carried out with Week (1, 2, 3) as within subjects variable and Training (Answer, Choice, Missing Addend condition) and Level (poor, average, good mathematical ability) as between subjects variables. The dependent variable was the percentage of errors in each week of practice. There was a main effect of Week, $F(2, 180) = 21.15$, $p < .001$, partial $\eta^2 = .19$ (with a Greenhouse-Geisser correction), namely errors decreased over the practice period. There was a main effect of Level, $F(2, 90) = 18.48$, $p < .001$, partial $\eta^2 = .29$, namely children with poor mathematical ability had lower scores than children with average ($p < .001$) and good mathematical ability ($p < .001$). Also, there was a main effect of Training, $F(2, 90) = 5.78$, $p < .01$, partial $\eta^2 = .11$. Post hoc Bonferroni tests showed that children in the Missing Addend condition made significantly more errors than children in the Answer condition ($p < .05$). Error rate in the Choice condition was not significantly different from the error rate in the Answer condition ($p = .45$) or from the rate in the Missing Addend condition ($p = .45$).

Table 4.3. Means and standard deviations of percentage of errors during practice for the three training conditions as a function of week of practice and children's mathematical ability.

	Week 1		Week 2		Week 3		Total	
Level	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Answer condition								
Poor	11.1	12.7	2.3	2.2	3.5	3.9	5.6	5.3
Average	4.0	4.8	3.2	2.4	2.7	2.3	3.3	2.1
Good	2.4	2.8	2.1	2.3	1.8	1.3	2.1	2.0
Total	5.9	8.4	2.7	2.3	2.7	2.7	3.8	3.5
Choice condition								
Poor	18.4	18.3	13.1	20.1	6.1	9.0	12.5	15.0
Average	3.2	3.2	4.0	6.2	6.0	7.7	4.4	5.1
Good	2.7	4.7	1.9	1.1	1.4	1.6	2.0	1.8
Total	7.5	12.3	6.1	12.2	4.9	7.3	6.2	9.6
Missing Addend condition								
Poor	40.6	34.3	15.9	16.2	8.1	7.9	21.5	13.7
Average	6.6	4.8	3.1	3.2	2.9	3.5	4.2	2.8
Good	8.2	5.3	4.1	2.7	2.7	2.4	5.0	2.2
Total	15.4	22.3	6.5	9.8	4.2	5.2	8.7	10.2

All interaction effects were significant, but we describe only the two interaction effects involving both Week and Training, because these are most relevant for the development over the practice weeks. There was a significant interaction between Week and Training, $F(4, 180) = 5.24$, $p < .01$, partial $\eta^2 = .10$ (with a Greenhouse-Geisser correction), indicating that the pattern of answered items over the weeks was different for the three training conditions. Within subjects contrasts showed a significant decrease in errors from Week 1 to Week 2 for children in the Answer ($p < .05$) and children in the Missing Addend condition ($p < .05$), but not for children in the Choice condition. From Week 2 to Week 3 there was no significant improvement for any of the training conditions. Over the whole practice period there was a significant difference in the improvement between children in the Choice condition (who improved from 7.5% errors in Week 1 to 4.9% errors in Week 3) and children in the Missing Addend condition (who improved from 15.4% errors in Week 1 to 4.2% errors in Week 3); posthoc Bonferroni tests showed a significant difference between the two groups on the difference between Week 1 and Week 3 ($p < .05$). There was no significant difference between the Answer condition and the other two training conditions.

The interaction between Week, Training, and Level was also significant, $F(8, 180) = 2.21$, $p < .05$, partial $\eta^2 = .09$ (with a Greenhouse-Geisser correction). The groups were too small to carry out detailed contrast analyses; therefore, only what is shown in Table 4.3 is described. However, because of the lack of statistical power, these observations should be treated with caution. In the Answer condition, the results for children with different ability in mathematics showed that the large decrease in errors from Week 1 to Week 2 could be attributed almost solely to children with poor mathematical ability. It is not surprising that children with poor ability improved most, because they made most errors to begin with. Nevertheless, the results confirm that the Answer method was effective for children with poor mathematical ability. For the Choice condition the results in Table 4.3 show that, whereas children with either poor or good mathematical ability made less errors in every week of the practice period, the percentage of errors actually increased for children with average mathematical ability. This suggests that the Choice method may be less effective for children with average mathematical ability. The results for children in the Missing Addend condition showed that, although the problems were initially difficult, all children learned much from practice. This was true even for the students with poor mathematical ability, that is, their performance was rather low in the first week of practice, but their scores improved much in Week 2 and Week 3.

4.3.3 Pretest, posttest and retention test

On the pretest children performed slightly better on the problems of Set B than on the problems of Set A, $t(102) = 2.21$, $p < .05$, Cohen's $d = .22$. Mean score for Set A was 12.6 problems ($SD = 8.2$) solved correctly in 2 minutes and mean score for Set B was 13.4 problems ($SD = 7.8$). This could be due either to Set A being slightly more difficult than Set B, or it could be an effect of the order in which the two sets were tested (on all moments of testing Set A was tested before Set B). To evaluate the possibility that the specific set of problems influenced the results of practice, Set was entered as a between subjects variable in the repeated measures ANOVA. Because Set did not interact with the Training condition, all analyses described hereafter do not include Set as a variable.

Mean scores for practiced and unpracticed items on the pretest, posttest, and retention test are presented in Table 4.4. A repeated measures ANOVA was carried out with Time (pretest, posttest, retention test) and Problem (practiced, unpracticed) as within subjects factors and Training (Answer, Choice, Missing Addend condition) as between subjects factor. The dependent variable was the number of items solved correctly within 2 minutes. A main effect of Time was found, $F(2, 200) = 61.90$, $p < .001$, partial $\eta^2 = .38$ (with a Greenhouse-Geisser correction), that is, performance increased with each test. Mean score was 13.0 ($SD = 7.8$) on the pretest, 16.9 ($SD = 9.0$) on the posttest, and 19.9 ($SD = 10.1$) on the retention test. There was also a main effect of Problem, $F(1, 100) = 10.26$, $p < .01$, partial $\eta^2 = .09$, namely significantly more practiced items ($M = 17.2$, $SD = 9.1$) were solved than unpracticed items ($M = 16.0$, $SD = 7.7$). Furthermore, there was a significant interaction effect of Time × Problem, $F(2, 200) = 7.20$, $p < .01$, partial $\eta^2 = .07$, Time × Training, $F(4, 200) = 3.05$, $p < .05$, partial $\eta^2 = .06$ (with a Greenhouse-Geisser correction), and Time × Problem × Training, $F(4, 200) = 2.69$, $p < .05$, partial $\eta^2 = .05$.

Table 4.4. Means and standard deviations of the number of items correct in the two paper and pencil two-minute tests at the pretest, posttest, and retention test (in the Answer format). Practiced and unpracticed items are presented separately for the three training conditions.

		Pretest		Posttest		Retention	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Practiced items							
Answer	35	13.3	8.8	20.9	11.9	21.0	11.9
Choice	39	12.2	7.8	18.8	9.9	20.3	11.3
Missing Addend	29	13.0	7.4	13.9	9.3	20.4	11.8
Unpracticed items							
Answer	35	13.4	8.6	16.5	9.1	19.9	9.6
Choice	39	13.4	8.2	16.4	7.3	18.4	9.2
Missing Addend	29	12.5	7.7	13.7	7.6	19.3	9.8

To gain an understanding of the three-way interaction, the scores on the posttest were compared to the scores on the pretest by carrying out a repeated measures ANOVA for each of the three training conditions. From pretest to posttest, there was a significant difference in gain between practiced and unpracticed items for children in the Answer condition, $F(1, 34) = 15.23$, $p < .001$, partial $\eta^2 = .31$, and for children in the Choice condition, $F(1, 38) = 10.45$, $p < .01$, partial $\eta^2 = .22$, but not for children in the Missing Addend condition ($p = .83$). Bonferroni tests showed that the gain on practiced items in the Missing Addend condition was significantly lower than the gain on practiced items in the Answer ($p < .01$) and in the Choice condition ($p < .05$). There was no significant difference between the gain on practiced items in the Answer condition and the gain on practiced items in the Choice condition, even though children in the Answer condition gained nearly one problem more on average. Furthermore, no significant differences between the training conditions were found in the gain on unpracticed items.

Long-term effects were evaluated by comparing the scores on the retention test to the scores on the pretest. From pretest to retention test, the three-way interaction between Time, Problem, and Training was no longer significant ($p = .33$). There was a main effect of Time, $F(1, 100) = 92.06$, $p < .001$, partial $\eta^2 = .48$, namely performance was better on the retention test than on the pretest. Furthermore, there was a significant interaction between Time and Problem, $F(1, 100) = 5.14$, $p < .05$, partial $\eta^2 = .05$, namely improvement was higher for practiced (mean gain 7.8, $SD = 8.6$) than for unpracticed items (mean gain 6.0, $SD = 7.6$). Mean score on the retention test was 20.6 ($SD = 11.5$) for practiced items and 19.2 ($SD = 9.5$) for unpracticed items. This means that a month after practice there was still a detectable effect of the training, but no significant differences between the training conditions were found.

Because it is possible that other methods of practice are effective for children with poor than for children with good mathematical ability, the analysis was repeated with both Training and Level (poor, average, good mathematical ability) as between subjects factors. No interaction effects between Level and the other variables were found (p value for Time × Problem × Training × Level was .46), suggesting that the pattern of results was the same for all children, irrespective of their aptitude in mathematics.

Finally, we examined the results of the two-minute tests of the posttest in all three presentation formats. Note that the choice problems and missing-addend problems were administered only at the posttest, so there were no pretest or retention-test results. Figure 4.2 shows the mean scores in all three presentation formats as a function of training condition. Practiced and unpracticed items are presented separately. Better scores on practiced compared to unpracticed items imply that item-specific learning has occurred as a result of practice. Considering first the results of the Choice test, it appears that there was an item-specific learning effect for all three training conditions. This observation was evaluated by means of a repeated measures ANOVA with Problem (practiced, unpracticed) as within subjects factor and Training (Answer, Choice, Missing Addend condition) as between subjects factor. A significant effect of Problem was found, $F(1, 100) = 20.49$, $p < .001$, partial $\eta^2 = .17$, that is, scores for practiced items were higher ($M = 20.8$, $SD = 11.2$) than scores for unpracticed items ($M = 18.2$, $SD = 9.8$). The effect of Training was not significant ($p = .64$) and no interaction between Problem and Training was found ($p = .17$). These findings suggest that children from all three conditions (including the Missing Addend condition) performed better on practiced compared to unpracticed items and that performance across the groups was similar. The result that children in the Choice condition did not show a significantly larger item-specific learning effect than the other practice conditions was somewhat unexpected, because we had anticipated that children who practiced in a format normally not used in education would perform better than other children on a test with items presented in the same format.

In contrast, the anticipated advantage for children who practiced in the respective condition was indeed present in the Missing Addend test. As can be seen from Figure 4.2, children in the Missing Addend condition showed an item-specific learning effect whereas children from the other practice conditions did not. Again, results were evaluated with repeated measures ANOVA. A significant main effect of Problem was found, $F(1, 100) = 10.63$, $p < .01$, partial $\eta^2 = .10$, namely scores for practiced items were higher ($M = 9.5$, $SD = 8.2$) than scores for unpracticed items ($M = 8.1$, $SD = 5.8$). There was also a main effect of Training, $F(2, 100) = 4.14$, $p < .05$, partial $\eta^2 = .08$, indicating that children from different training conditions scored differently. Mean score for the Answer condition was 7.4 ($SD = 5.5$), mean score for the Choice condition was 7.9 ($SD = 6.2$), and mean score for the Missing Addend condition was 11.7 ($SD = 7.5$). Finally, a significant interaction effect between Problem and Training was found, $F(2, 100) = 10.18$, $p < .001$, partial $\eta^2 = .17$. To explore the interaction effect another repeated measures ANOVA (with Problem as between subjects factor) was carried out for each of the three training conditions. A significant difference in score between practiced and unpracticed items was found only for children in the Missing Addend condition, $F(1, 28) = 13.98$, $p < .01$, partial $\eta^2 = .33$, namely scores for practiced items (M =14.0, SD = 10.1) were higher than scores for unpracticed items (M = 9.3, SD = 5.6). Furthermore, Bonferroni tests showed that the score on practiced items was significantly higher for children who practiced in the Missing Addend condition compared to children who practiced in the Answer ($p < .01$) or in the Choice condition ($p < .01$). No significant differences between the training conditions were found on the unpracticed items.

Figure 4.2. Number of correct items in the two-minute tests at the posttest. Children were tested on practiced and unpracticed items in the three practice formats: Answer, Choice and Missing Addend. For each format, the mean scores are presented separately for the three training conditions.

On the whole, children performed quite poorly on the Missing Addend test, with an average score of 8.8 problems correct in 2 minutes. There were 10 children who did not manage to answer any of the problems correctly: three children in the Answer condition, five in the Choice condition, and two in the Missing Addend condition. The two children in the Missing Addend condition had poor mathematical ability and did not solve many problems (69 and 74 problems, respectively) during the practice period. Nevertheless, during practice they did solve most of the problems correctly (70% and 73% correct, respectively).

4.4 Discussion

In this study we evaluated the effectiveness of three different methods of practice to improve children’s performance on simple addition problems: (a) writing down the answer (Answer method), (b) choosing the correct answer out of two alternative answers (Choice method), and (c) filling in the missing second addend (Missing Addend method). Grade 1 children practiced for three weeks, three times a week for 10 minutes, on a small set of problems. Performance on practiced and unpracticed items was measured before (pretest) and after the practice period (posttest and retention test) with two paper and pencil two-minute tests, presented in the Answer format. In addition, on the posttest children solved the problems not only in the Answer format, but also in the other two formats.

4.4.1 Practice results

It was expected that the Choice method would be easier than the Answer method, whereas the Missing Addend method would be more difficult than the Answer method. This appeared to be the case, although not all of our hypotheses were confirmed. During practice, the Choice method was significantly faster than the Answer method and, therefore, Hypothesis 1a was confirmed. However, there was no significant difference between errors made in the Choice condition and errors made in the Answer condition, thus falsifying Hypothesis 1b; had there been a difference, then the direction of the effect would be contrary to expectations. Furthermore, children in the Missing Addend condition made more errors than children in the Answer condition and, therefore, Hypothesis 2b was confirmed. However, this method was not significantly slower than the Answer method, thus falsifying Hypothesis 2a.

To gain more insight in the process of learning, the development of speed and accuracy over the three practice weeks was also studied. Children in the Answer condition improved a little on speed, but the main improvement was on accuracy. This was especially true for the children with poor mathematical ability, who showed an enormous improvement on accuracy, but children with average and good mathematical ability improved as well. The Choice method appeared to be demotivating for children with average mathematical ability, because they solved fewer problems and made more errors in each week of the practice period. Maybe it was also a little demotivating for children with good mathematical ability, because they solved fewer problems at the end than at the beginning of the practice period, but they still improved on accuracy. However, for children with poor mathematical ability the Choice method worked out fine, because they improved both on accuracy and on speed. The missing-addend problems were especially difficult for children with poor mathematical ability. However, performance for all children in the Missing Addend condition improved during the practice period: children with poor mathematical ability improved on accuracy, and children with average and good mathematical ability improved both on accuracy and on speed.

4.4.2 Pretest, posttest, and retention test results

The results of the pretest and posttest showed that both the Answer and the Choice training condition were effective for learning addition facts: improvement for practiced items was higher than for unpracticed items. Both methods of practice were equally effective. This means that there was a direct learning effect for children in the Answer condition and a positive transfer effect for children in the Choice condition, and, therefore, Hypothesis 3b was confirmed. The Missing Addend training condition was not effective: scores hardly improved and improvement was not higher for practiced compared to unpracticed items. This means that there was no transfer from practicing missing-addend problems to regular addition problems, and, therefore, Hypothesis 4b was also confirmed. Results were the same for children with poor, average, and good mathematical ability; consequently, Hypotheses 5a and 5b were falsified. When the Answer test was repeated one month later (retention test), there was still a small learning effect – improvement compared to the pretest was higher for practiced than for unpracticed items – but there were no differences between the three methods of practice.

The results for children in the Answer condition showed that frequent repetition of a small set of problems is an effective method to improve performance on simple addition. The average score for practiced problems improved from 13 correct items (in 2 minutes) at the pretest to 21 correct items at the posttest. This means that if no mistakes were made, each problem was solved in about 5.7 seconds at the posttest, compared to 9.2 seconds at the pretest. Improvement in speed on solving problems may be caused by faster procedural skills, a transition from counting to retrieval strategies, faster retrieval, or most likely a mixture of these factors (Geary, Brown, & Samaranayake, 1991; Goldman, Mertz, & Pellegrino, 1989; Imbo & Vandierendonck, 2008a, 2008b). The most striking result, however, was that choosing from alternative answers was equally effective as answering addition problems in the conventional way.

Improving on answering addition problems after practicing with the Choice method may seem puzzling at first sight, because choosing the correct answer comes across as a “mindless” procedure. However, according to network theories on the storage of arithmetic facts it is important that an arithmetic problem is regularly associated with the correct solution (Ashcraft, 1992; Campbell, 1995; McCloskey et al., 1991; Siegler, 1988; Verguts & Fias, 2005). The results of the present study provide support for these network models: strengthening the association between a problem and the correct answer appears to be an effective method to improve children’s performance on simple addition problems. The success of the Choice method is probably partly due to the fact that it is fast, compared to the conventional way of answering addition problems. On average, children in the Choice condition made each problem 39 times during practice, whereas children in the Answer condition made each problem 28 times. According to network theories, frequent presentation of problems is an important factor for performance in simple arithmetics. Because children in the Choice condition made many more problems than children in the Answer condition and nonetheless did not show a higher improvement, we may conclude that choosing between alternative answers is less effective than the conventional method as measured by number of items made. An alternative explanation is that for both conditions the amount of practice was “sufficient”. Such a hypothesis is warranted by the observation that scores for practiced items hardly improved in the Answer and Choice condition during retention period, whereas scores in the Missing Addend condition improved from 14 to 20 items. Settling the matter of which method is most effective as measured by number of problems made would require all children to make the same number of problems. However, for educational implications this issue is not particularly relevant: the results of the present study show that with practice time being equal, choosing between answers is just as effective as writing down the answer.

There was absolutely no transfer from practicing missing-addend problems to regular addition problems. This finding is in accordance with the Identical Elements (IE) model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996; Rickard et al., 1994), which predicts that when the numerical elements of a test problem do not match exactly with those of a practice problem, there will be no positive transfer. Judging from the current findings, the IE model seems an appropriate model to explain the results of children in the Missing Addend condition. There is also a second prediction that we could make from the IE model. In general, positive transfer would be expected from practice problems (e.g., 3 + 2 = …) to test problems with an operand order change (2 + 3 = …). Because 6 of the 13 problems in the two sets (Set A and Set B) had their commutative counterpart in the other set, we would expect that children who improved on the practiced problems, also improved a little on the unpracticed problems. If we analyze the posttest results, this seemed to be what happened. Children in the Answer and Choice conditions improved not only on practiced problems, but also on unpracticed problems. However, no significant differences were found between the three conditions on unpracticed items, that is, the children in the Answer and Choice conditions did not score higher on unpracticed problems than children in the Missing Addend condition. Thus, although the scores on unpracticed items suggest a general learning effect for children in the Answer and Choice conditions, next to the item-specific effect of having higher scores for practiced compared to unpracticed items, this effect was not significant.

The training had a lasting effect on the children who participated in the study; after a month (retention test) there was still a detectable item-specific learning effect. This effect was marginal but statistically significant. Finding only a small learning effect on the retention test is not surprising, because the retention period consisted of about three school weeks (and one week of holidays) during which simple addition problems were also addressed almost every day. No effect of practice condition appeared on the retention test. The fact that the learning effect was very small probably precludes an interaction effect with practice condition statistically. Therefore it is difficult to interpret these retention effects other than by concluding that a small, but enduring advantage had been established for the items that have been practiced. In all, for educational practice it is valuable to experience that investing 10 minutes on 10 different days has a lasting effect on learning.

4.4.3 Results in the Choice and Missing Addend formats of the posttest

On the posttest, not only the Answer format was presented, but also the other two formats. This enabled us to look at format-specific effects. The choice two-minute tests, administered directly after the practice period, showed the presence of a learning effect for children who practiced with the Choice method, thus confirming Hypothesis 3a. Furthermore, children in the other two practice conditions showed a transfer effect that did not differ from the learning effect in the Choice condition. The missing-addend two-minute tests showed the presence of a learning effect for children who practiced with the missing-addend method. This finding confirmed Hypothesis 4a. However, in contrast with the choice tests, there was no transfer effect in the other practice conditions.

The results show that children who practiced with the missing-addend method, gained efficiency in solving the 13 problems they practiced. As mentioned before, there was no transfer of this knowledge to problems that were presented in the Answer format. Remarkably, on the choice two-minute tests no differences in learning effect arised between children in the Missing Addend condition and children from other practice conditions. All children scored higher on practiced compared to unpracticed items and there were no significant differences between practice conditions in their scores on the choice two-minute tests. This implies, first of all, that children in the Missing Addend condition did gain some knowledge about simple addition facts, although this was not revealed by the answer two-minute tests.

Another important conclusion is that production and verification tests measure mathematical ability at a different level. This has been put forward by others (Dagenbach & McCloskey, 1992; Zbrodoff & Logan, 1990), but it is important to be aware of this when a method is selected to measure children’s mathematical performance. In order to measure performance in production one should choose a production test and not a verification test. However, the merit of a choice test is that it shows the passive knowledge present in children, which makes it a sensitive tool to reveal learning effects.

4.4.4 Educational implications

First of all, the present study showed that children can make significant progress in a short amount of time if they practice simple addition problems. Frequent repetition is an important factor in this process, and it seems that practicing a small set of problems over and over again is very effective. Of course, practice can only be effective if a child has a basic idea on how to solve the problems. Therefore, the timing is important. In the present study, children were familiar with the addition problems they solved, although they had not yet practiced them frequently. However, a child that lacks important conceptual knowledge about addition probably will not gain very much from practice. Therefore, frequent repetition as used in the Answer or Choice method will probably not be effective for children who are unable to solve the problems on their own, for instance children with serious difficulties in learning mathematics. Children with mathematical difficulties are not only delayed in the development of computational skills, they also have considerable problems with arithmetic fact retrieval (Geary, 1993). These fact retrieval difficulties are persistent throughout primary school and extensive drilling often does not seem to substantially improve performance (Geary, 1993). Thus, the Answer and Choice method might not be effective for these children. Nonetheless, the present results show that relatively normal developing children with below average performance in mathematics benefit from frequent repetition.

Second, the study shows that choosing from two alternative answers as a method of practicing addition problems is almost as effective as writing down the answer. An advantage of this method seems to be that problems are solved very fast, which enables frequent repetition. However, a disadvantage is the observed "sloppiness" of the children who practiced with the Choice method; they made more errors during practice than children who wrote down the answers. Higher proportion of errors during practice in the Choice condition is likely due to the very nature of the task: making a choice between two answers. Furthermore, there is of course a risk that presenting children with incorrect answers strengthens the wrong associations. To avoid this effect in the current study, three different incorrect answers were used for each problem. The results in the Choice condition do not indicate that children were confused by the incorrect answers they had seen, but to know this for sure one would have to look at production errors after practicing with the Choice method. With respect to different levels of ability in mathematics, the learning effects as measured on the tests were similar for children with poor, average, and good mathematical ability. However, the findings in the practice data suggest that the Choice method is most promising for children with poor mathematical ability, who improved both in speed and in accuracy during practice. Practicing with this method could be a useful alternative of practicing addition problems in the traditional way.

Finally, as expected, there was no transfer of knowledge from practicing with problems in the Missing Addend format to problems presented in the Answer format. This means that this method cannot be used to obtain automaticity in addition facts. Of course, missing-addend problems can still be useful as a means to contribute to children's conceptual knowledge about addition. In Dutch education, addition problems with a missing addend are sometimes used to enhance conceptual understanding (e.g., "Someone spilt ink on the second digit of the problem, can you solve 3 + … = 5?"). Solving such a problem requires sophisticated knowledge of the structure of an addition problem. Missing-addend problems can also be used to represent and solve simple word-problems. In addition, they can be a means to attain understanding of the inverse relationship between addition and subtraction (Baroody, 1999a). We believe, however, that missing-addend problems should be introduced when children have more experience with solving addition problems. For instance, we noticed that a negligible number of children in the Missing Addend condition made errors in the answer two-minute tests on the posttest because they mistakenly subtracted the addends. This suggests that the Missing Addend method might confuse the children in this phase of mathematical development.

4.4.5 Conclusion

In summary, the present results show that choosing from alternative answers is effective as a method to obtain automaticity in addition facts. From a theoretical point of view, the present study clearly provides support for network models on learning arithmetic facts (Ashcraft, 1992; Campbell, 1995; McCloskey et al., 1991; Siegler, 1988; Verguts & Fias, 2005). If a child succeeds in finding the answer to a problem on different occasions, the association between problem and solution becomes stronger. Producing the answer is of course an effective way of strengthening the association between problem and answer, as this is used all the time in mathematics education. However, selecting the correct answer from alternatives also strengthens the association in a very effective way. With respect to problems with a missing addend, the results suggest that the Identical Elements model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996; Rickard et al., 1994) is a useful theory to explain transfer effects in children on simple addition.

4.5 Acknowledgments

We are grateful to all the students and teachers who participated in this study.

Appendix A

Problems of Set A, Set B, and fillers. The listed incorrect alternatives were used in the Choice condition.

Set A				Set B				Fillers		
	Incorrect				Incorrect				Incorrect	
Problems	alternatives			Problems	alternatives			Problems	alternatives	
3 + 2 = 5	4	7	10	2 + 3 = 5	4	7	8	1 + 2 = 3	4	6
2 + 4 = 6	3	5	8	1 + 5 = 6	3	4	7	2 + 1 = 3	5	6
3 + 3 = 6	4	5	8	4 + 2 = 6	5	8	9	1 + 3 = 4	2	5
1 + 6 = 7	3	5	8	3 + 4 = 7	5	6	9	2 + 2 = 4	3	6
5 + 2 = 7	6	9	10	4 + 3 = 7	5	6	10	3 + 1 = 4	5	7
1 + 7 = 8	5	9	10	2 + 6 = 8	3	7	10	1 + 4 = 5	3	6
3 + 5 = 8	4	6	9	5 + 3 = 8	6	9	10	4 + 1 = 5	6	8
4 + 5 = 9	6	7	8	1 + 8 = 9	5	6	7	5 + 1 = 6	7	8
5 + 4 = 9	6	7	10	3 + 6 = 9	5	7	8	2 + 5 = 7	4	8
6 + 3 = 9	7	8	10	7 + 2 = 9	8	10		6 + 1 = 7	8	9
2 + 8 = 10	6	7	9	3 + 7 = 10	5	6	9	4 + 4 = 8	5	9
4 + 6 = 10	5	8	9	5 + 5 = 10	6	8	9	6 + 2 = 8	7	10
7 + 3 = 10	8	9		6 + 4 = 10	7	8	9	7 + 1 = 8	9	10
								2 + 7 = 9	6	8
								8 + 1 = 9	7	10
								1 + 9 = 10	5	7
								8 + 2 = 10	8	9
								9 + 1 = 10	6	7

5

Transfer effects in children's addition and multiplication

Predictions of the Identical Elements (IE) model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996) about transfer between number facts were tested in primary school children. The aim of the study was to test whether the IE model, constructed to explain adult performance, also applies to children. The IE model predicts no positive transfer when the numerical elements of a test problem do not match exactly with those of a practice problem. In two experiments, children practiced addition or multiplication problems. A positive transfer effect was found for problems with an operand order change; improvement was just as high as for practiced problems. No transfer effect was found for problems with one of the operands increased with one unit; improvement for these problems did not differ from problems unrelated to the practice problems. Analogous results were found for addition and multiplication, suggesting that storage and retrieval processes in both domains are highly similar in children. The findings corroborate the hypothesis that the IE model applies also to children who are starting to master simple arithmetic.

Submitted as: Van Galen, M. S., & Reitsma, P. Transfer effects in children's addition and multiplication.

5.1 Introduction

For most adults, the fastest strategy to solve simple addition and multiplication problems is to retrieve the answer from long-term memory. Usage of other strategies has also been reported, for instance making use of derived facts (e.g., 3 + 4 = [3 + 3] + 1 or 9 × 8 = [10 × 8] - 8) or counting, but retrieval is the most dominant strategy (Campbell & Austin, 2002; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). It is generally assumed that the storage of arithmetic facts in long-term memory is organized as an associative network. Various network models have been proposed for arithmetic fact retrieval (Ashcraft, 1992; Campbell, 1995; McCloskey, Harley, & Sokol, 1991; Rickard, 2005; Siegler, 1988; Verguts & Fias, 2005). The models have different characteristics, but they have in common that an arithmetic problem is connected with several possible solutions. When the problem is solved correctly on repeated occasions, the association between the problem and the correct answer becomes stronger, i.e., the probability for correct retrieval increases. There is much empirical evidence for the storage of arithmetic facts in an associative network, coming from studies on production errors (e.g., Campbell, 1994) and on interference and priming effects (e.g., Campbell, 1987; Campbell & Clark, 1989).

Arithmetic fact retrieval is well studied in adults, but much less research has been devoted to developmental aspects of the storage and retrieval of arithmetic facts in children. Through extensive practice with addition problems, children in primary school move on from initial counting procedures to faster memory-based strategies (Geary, Bow-Thomas, Liu, & Siegler, 1996; Goldman, Mertz, & Pellegrino, 1989; Siegler, 1987). A similar development occurs in multiplication; with increasing experience, strategies like repeated addition (e.g., 3 × 4 = 4 + 4 + 4) or counting sequences (e.g., 3 × 5 = 5, 10, 15) are replaced by retrieval (Cooney, Swanson, & Ladd, 1988; Mabbott & Bisanz, 2003; Lemaire & Siegler, 1995). Throughout primary school, children develop an associative network for arithmetic facts. Lemaire, Barrett, Fayol, and Abdi (1994) found evidence that children as young as seven activate addition facts automatically. When asked to indicate whether a probe had been present in a previously viewed number pair, children from different ages took more time to reject distractors equal to the sum of the pair than unrelated numbers. However, the interference effect depended on the size of the numbers in the pair and the age of the child. Seven-year-olds showed an interference effect only when both integers in the original pair were 5 or smaller, whereas eight-year-olds also showed an interference effects when one integer was 5 or smaller and the other was between 6 and 9. Nine- and ten-year-olds showed an interference effect on all presented problems. These findings suggest that children gradually expand their associative network for arithmetic facts.

In the present study, we focus on children who are in the beginning of this process. Our aim was to answer two questions about the organization of arithmetic facts in memory. The first question is whether commutative problems are stored separately: is there only one representation for both 5 + 2 and 2 + 5 or are the problems represented by two distinct facts in memory? If there is only one representation for both orders, it is expected that practicing 5 + 2 not only leads to an improved performance on 5 + 2, but also to an improved performance on 2 + 5. The second question is whether connections exist between arithmetic facts that likely are neighbors in the associative network; for instance, do children realize that they can calculate 5 + 3 if they know the answer to 5 + 2? In two experiments, children repeatedly practiced a small set of addition or multiplication problems. After practice, participants were tested not only on the practiced problems, but also on several unpracticed problems. Transfer effects were evaluated for two different categories of related problems. 1) Problems with an order change; e.g., 6 × 3 was practiced and 3 × 6 was presented on the test. 2) Problems with one of the operands increased with one unit; e.g., 6 × 3 was practiced and 7 × 3 was presented on the test. The present question is whether children's practice on one set of problems leads to an improved performance on another set with related problems.

According to the Identical Elements (IE) model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996; Rickard, Healy, & Bourne, 1994), there is a single long-term memory node for problems consisting of the same numerical elements (i.e., operands and answer), regardless of operand order. The IE model predicts that when the numerical elements of a test problem do not match exactly with those of a practice problem, there will be no positive transfer. This model was tested in adults with multiplication and division. Multiplication problems with an operand order change (e.g., 4 × 6 = _ practiced, 6 × 4 = _ on the test) were answered almost as fast as test problems that were identical to the practiced problems (Rickard & Bourne, 1996). The same was true for problems with a symbol change that consisted of the same numerical elements (e.g., _ : 4 = 6 practiced, _ = 4 × 6 on the test) (Rickard et al., 1994). However, when the numerical elements of a test problem and a practice problem differed from each other, there was no positive transfer. Similar findings were obtained in a priming study on addition and subtraction by Campbell, Fuchs-Lacelle, and Phenix (2006).

The identical elements model was constructed as a theory to explain adult performance on arithmetic fact retrieval. In other words, it describes a situation in which the associative network in memory is more or less established. Is the IE model useful to describe transfer effects also in children, who are still in the process of mastering arithmetic facts? Baroody (1999b) found that children use relational knowledge, such as the commutativity principle, to find the answer to multiplication problems. Although commuted combinations like 8 × 3 and 3 × 8 might initially be stored as distinct facts in long-term memory, Baroody hypothesized that the discovery of the commutativity principle induces a reorganization in the mental representation of multiplication facts, so that eventually only one representation is stored. There is indeed some evidence consistent with this hypothesis (Butterworth, Marchesini, & Girelli, 2003). Italian children of different ages were tested on a set of simple multiplication problems. In contrast to most other countries in Europe and to the United States, children in

Italy learn multiplication facts like 2 × 6 before they learn 6 × 2, because the 2 × *n* table (2 × 1, 2 × 2, 2 × 3, …) is introduced before the 6 × *n* table. If frequency effects shape the associative network for multiplication facts (Ashcraft & Christy, 1995; Geary, 1996; Siegler, 1988), it is expected that Italian children solve 2 × 6 faster than 6 × 2, but instead a time advantage was found on problems with the largest operand first (6 × 2). However, this effect was significant only in nine- and ten-year-olds, and not in eight-year-olds, thus suggesting that multiplication facts are reorganized with growing mathematical skills. Since the older children were systematically faster in responding to problems with the largest operand first, Butterworth et al. assumed that the representation of multiplication facts is reduced in memory, in the sense that only half of the problems is represented. To solve a problem like 2 × 6, a transformation to the standard form 6 × 2 is required, resulting in a slightly larger reaction time. However, it is also possible that both forms are separately represented in memory and that the representation with the largest operand first is easier to access than the commuted representation.

There are of course more connections between different arithmetic facts than just the commutative relations. It is, for instance, highly plausible that problems from the same multiplication table are interconnected in long-term memory. This hypothesis is supported by the observation that most production errors in adults' multiplication are table related (e.g., Campbell, 1994); an example of such an error is 6 × 4 = 28 instead of 24. Furthermore, table related errors are characterized by a distance effect: responses are more likely to be correct answers for problems nearby in the multiplication tables than for problems further away. For example, 6 × 4 = 28 is a more common error than 6 × 4 = 32. Galfano, Rusconi, and Umiltà (2003) showed that presentation of two numbers (e.g., 6 and 4) automatically activates not only the multiplication problem 6 × 4 and its correct response 24, but also the immediate neighbors of that multiplication problem (5 × 4 = 20, 7 × 4 = 28, 6 × 3 = 18, 6 × 5 = 30). Galfano et al.'s findings suggest that, in adults, multiplication facts are stored in a highly related network in which activation spreads from the product node to adjacent nodes.

If children become aware of the relations that exist between different arithmetic facts, they might use their knowledge of specific arithmetic problems to solve related problems. In the present study, we selected addition and multiplication problems that were related to the practice problems in the sense that one of the operands was increased with one unit. This way, children could benefit from the fact that they had practiced a certain problem when asked to solve a related problem. For instance, the problem 6 + 4 might be solved as [6 + 3 = 9] + 1 if children learned 6 + 3 during practice. And 8 × 4 might be solved as [7 × 4 = 28] + 4 if children learned 7 × 4 during practice. We assume that a procedural strategy is used when improvement on related problems is higher than improvement on unrelated problems, but not as high as the improvement on the practiced problems. Although in relatively skilled adults transfer probably does not include procedural strategies (Rickard & Bourne, 1996; Rickard et al., 1994), the participants in the present study are children who are still struggling to find the answer to simple arithmetic problems. For them, procedural strategies may be much more beneficial and therefore a positive transfer effect might be found for related problems. To be sure, transfer effects are not necessarily positive. In a study of Campbell (1987), adults practiced a subset of multiplication problems. After the practice period, performance on unpracticed problems was worse (more errors and slower reaction times), compared to performance on a test prior to the practice period. Apparently, the practiced problems (e.g., $7 \times 8 = 56$) interfered with the unpracticed problems (e.g., 56 was answered to 7×9). Therefore, it is also possible that a negative transfer effect is found for the related problems in the present study.

To sum up, in the present study transfer effects in addition and multiplication are studied in children. In the first experiment children in Grade 1 (mean age 7.2 years) practiced a subset of addition problems, and in the second experiment children in Grade 3 (mean age 8.6 years) practiced a subset of multiplication problems. In the two experiments different measures for performance were used: reaction times in the addition experiment and accuracy (within a certain time limit) in the multiplication experiment. The selection of an appropriate dependent variable is related to the level of skill in the participants. In the addition experiment participants were familiar with the problems, so in general they were able to give the correct answer. If the majority of the answers is correct, reaction times are suitable to measure performance. In contrast, the multiplication experiment included not only familiar problems but also problems that participants had no experience with. For this reason, reaction times were not adequate as a measure for performance in the multiplication experiment and accuracy was chosen as the dependent variable. Before and after practice, participants were tested on the practice problems, on problems with a change in operand order (e.g., $6 + 3$ practiced, $3 + 6$ on the test; 7×4 practiced, 4×7 on the test), on related problems (e.g., $6 + 3$ practiced, $6 + 4$ on the test; 7×4 practiced, 8×4 on the test), and on unrelated problems. If the identical elements model for arithmetic fact representation applies to children who are starting to master simple arithmetic, there will be a transfer of the knowledge gained during the practice period to the problems with an order change, but transfer is not expected for related problems. If however a transfer effect is found for related problems, this suggests that children used procedural strategies. The unrelated problems served as a control for general speed up, which could be caused by familiarity with the testing procedure or by faster procedural skills. The frequencies of the practice problems were varied; larger practice effects and transfer effects were expected on problems with a higher frequency. Results are analyzed separately for children with poor and with good mathematical skills, because learning effects might be different for the two groups. If a transfer effect is found for the related problems, we expect children with good mathematical skills to improve more than children with poor mathematical skills, because children with good mathematical skills probably have a better understanding of the relations that exist between different number facts.

5.2 Experiment 1: Addition

5.2.1 Method

5.2.1.1 Participants

The participants were 39 Grade 1 students (24 girls, 15 boys) from two Dutch schools for primary education. Mean age of the children was 7.2 years (*SD* = 0.29). The children were tested about one month before the end of the school year. They were familiar with the difficulty level of the problems used in this study (addition up to 10).

5.2.1.2 Material and procedure

The participants practiced for two weeks on a set of 5 addition problems with answers up to ten. Before and after the practice period they were tested both on the practiced problems and on 15 additional unpracticed problems. Practiced and unpracticed problems are listed in Table 5.1. The unpracticed problems fell into three categories: 1) commutative problems, 2) related problems, and 3) unrelated problems. In the commutative problems the first and second addends of the practice problems were switched. In the related problems the second addend of the practice problems was increased with one unit. The unrelated problems were unrelated to the practice problems in terms of the previous two categories. Of course, addition problems with answers up to ten are all related to each other in some way. Therefore, the difference between related and unrelated problems should be considered as a matter of degree rather than an absolute difference.

Table 5.1. Addition problems presented on the pretest and posttest. During the practice period, children practiced only the practice problems. Items '5 + 2' and '6 + 3' were high in frequency in version A of the training booklet; items '3 + 4' and '7 + 2' were high in frequency in version B.

Practice problems	Commutative problems	Related problems	Unrelated problems
2 + 3 = 5	3 + 2 = 5	2 + 4 = 6	4 + 2 = 6
5 + 2 = 7	2 + 5 = 7	5 + 3 = 8	2 + 6 = 8
3 + 4 = 7	4 + 3 = 7	3 + 5 = 8	5 + 4 = 9
6 + 3 = 9	3 + 6 = 9	6 + 4 = 10	3 + 7 = 10
7 + 2 = 9	2 + 7 = 9	7 + 3 = 10	4 + 6 = 10

The pretest and posttest were conducted on a laptop with a 15 inch screen and a display resolution of 800 × 600 pixels, with a separate response button. Stimuli were presented in Arial, font size 72. The experimenter told the children they should press the response button as soon as they knew the answer to the presented problem and that they were to utter the answer at the same time. There was no time limit for answering, but the instruction emphasized both accuracy and speed.

Participants were instructed to keep their hand close to the response button. The test consisted of three instruction items and two blocks with 20 experimental items; all problems were presented twice. Within an experimental block the 5 practiced and 15 unpracticed items were presented in a pseudo-random order: associated items (e.g., practice item '2 + 3', commutative item '3 + 2', and related item '2 + 4') were not presented consecutively. Problems from the same category were distributed over trials: the unrelated items were never presented directly after each other and from the other three categories there were at most two problems from the same category in consecutive trials. Finally, in consecutive trials the same number was used at most two times on the same position in the problem (first addend, second addend, outcome). A trial started with an empty screen (1 second), then an asterisk appeared as a fixation point in the middle of the screen (500 ms), then the screen was empty again (500 ms), and finally the addition problem was presented in the middle of the screen. As soon as the child pressed the response button the problem disappeared and the experimenter entered the answer of the child on a separate numeric keypad. To explore whether participants were reliable in their responses, a second person observed the registration of responses during the posttest. In about 5 % of the trials the utterance and button press were not simultaneous – in most of these cases the button press preceded the answer. However, the time between pressing and answering was generally very short, less than half a second. Because deviations were small and varied over items, we did not expect these differences to systematically influence the means. Pretest and posttest were identical except for the order of the items. On average, children needed about 7.0 minutes to finish the pretest and about 5.5 minutes to finish the posttest.

Training material were A4 size booklets with 1050 addition problems, comprised of 105 blocks with 10 problems. Two different versions of the training booklets were randomly assigned to the children. In version A problems '5 + 2' and '6 + 3' were listed twice as frequent as the other three problems and in version B '3 + 4' and '7 + 2' were listed in a double frequency. A block of 10 problems consisted of 7 practice items (5 items + 2 items repeated) and 3 filler items. Filler items were addition and subtraction problems with 1 or 0 ('x + 1', '1 + x', 'x – 1', 'x + 0', 'x – 0'), selected to have a variety of outcome numbers in the training booklets. The order of the problems was pseudo-random: there were never two identical problems in succession, there were never two filler items in succession, and in consecutive trials the same number was used at most twice on the same position in the problem (first addend, second addend, outcome). There were 50 problems on each page of the booklet, presented over two columns in groups of five.

All children in a school class participated. Every child received a personal booklet for practice. For two weeks, they practiced three times a week with their own teacher; each session lasted for 10 minutes. At the end of the week the booklets were corrected by the experimenter who made a mark if children made a mistake or forgot to fill in the answer. In the beginning of the second week, children studied their mistakes and tried to correct them before they carried on with the practice sessions. In the third week, children corrected their errors once again but did not practice. The posttest was one day after the final correction session. For each practice session 175 problems were listed in the booklet; this was the maximum number of problems that children could make in one session. The teacher instructed the children how to do the exercises: the problems had to be solved from top to bottom, before turning to the next column. It was not allowed to skip a problem. They were encouraged to solve the problems as accurately as possible.

5.2.1.3 Analysis

Originally, 44 children participated. Five children were excluded from the data set because they had too many invalid reaction times during pretest and posttest – they gave an incorrect answer or the RT was categorized as an outlier. After RTs of incorrect answers were removed, for each participant the mean RT and standard deviation was calculated. Then, an RT was categorized as an outlier when it deviated more than 2 standard deviations from the mean RT of the participant. In addition, 4 reaction times under 300 ms were also categorized as outliers. All outliers were removed. After this, the mean RT was calculated for every problem in the test. Because the problems were presented twice, the mean RT for a problem was based on two RTs. If there was only one valid RT, this RT was taken. If both RTs were invalid, the problem had a missing value. Children with more than 10 % missing values (4 out of 40 problems from the pretest and posttest) were excluded from the data set – these were the five children that were mentioned before. The remaining 39 children had 8.4 % invalid RTs (3.9 % incorrect answers and 4.5 % outliers), resulting in 1.86 % problems with a missing value for the mean RT. If a problem had a missing value on one of the tests, the RT for that problem was also removed from the other test, resulting in 3.7 % missing problems in total. Finally, the mean RT was calculated for each of the four categories (practice, commutative, related, and unrelated problems) in the pretest and posttest.

The two schools appeared to differ in enforcing the 10 minutes practice time, with the effect that children in school 1 made all 1050 problems in the training booklets and children in school 2 made on average 803 problems including fillers. This should not necessarily affect the results, because all four problem categories were tested within subjects, but we will include ‘school’ as a between-subjects factor in the analyses to control for possible interaction effects. With respect to errors made during practice: children made very few errors, mean percentage was 1.2 %. Performance improved during the practice period; in the first session the mean percentage of errors was 3.0 % and in the last session it was 0.7 %. The two schools did not significantly differ on percentage of errors made during practice.

5.2.2 Results

To evaluate if there was an improvement on any of the problem categories after practice, a repeated measures Anova was carried out with Time (pretest, posttest) and Problem (practice, commutative, related, unrelated) as within-subjects variables. The dependent variable was mean RT. Initially, School and Version (training booklet A or B) were included as between-subjects variables. The results showed no main effect for School or Version and no significant interaction effects with School. There was one significant interaction effect with Version: Problem × Version ($p < .05$). This means that children from version A scored differently on the four problem categories than children from version B. However, because there was no significant interaction effect between Time, Problem, and Version, we decided to leave both School and Version out of the final analysis.

The mean reaction times for the four problem categories on pretest and posttest are presented in Figure 5.1. The reaction times on the pretest suggest that some categories were more difficult than other. For instance, the related problems were answered more quickly than the unrelated problems (mean difference 127 ms). To see whether the four problem categories differed statistically from each other on the pretest, a repeated measures Anova was carried out with Problem (practice, commutative, related, unrelated) as within-subjects variable and RT on the pretest as dependent variable. No significant differences were revealed. The repeated measures Anova, with Time and Problem as within-subjects variables, showed a main effect of Time ($F(1, 38) = 31.17$, $p < .001$, $\eta_p^2 = .45$): reaction times were faster on the posttest ($M = 2549$ ms) than on the pretest ($M = 2960$ ms). There was also a main effect of Problem ($F(3, 114) = 5.12$, $p < .01$, $\eta_p^2 = .12$) and a significant interaction effect between Time and Problem ($F(3, 114) = 5.60$, $p < .01$, $\eta_p^2 = .13$, with a Greenhouse-Geisser correction). From Figure 5.1 it can be seen that reaction times for all four problem categories were faster on the posttest compared to the pretest. Within-subjects contrasts showed that the gain in RT for the commutative problems ($M = 553$ ms) did not significantly differ from the gain in RT for the practice problems ($M = 658$ ms). Furthermore, the gain in RT for the related problems ($M = 283$ ms) did not significantly differ from the gain in RT for the unrelated problems ($M = 152$ ms). All other contrasts were significant: the gain on practiced problems was different from related ($p < .01$) and unrelated problems ($p < .01$), and the gain on commutative problems was also different from related ($p < .05$) and unrelated problems ($p < .05$).

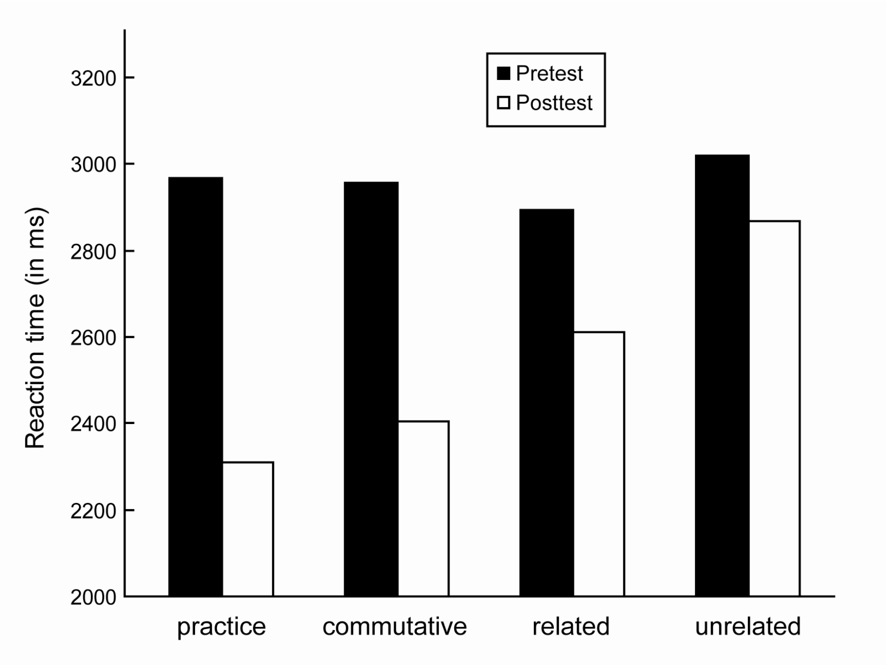

Figure 5.1. Addition problems in Grade 1. Mean reaction times (in ms) on the pretest and posttest for practice, commutative, related, and unrelated problems.

To evaluate frequency effects, separate mean reaction times were calculated for practice problems with a double frequency and practice problems with a regular frequency[1]. A repeated measures Anova was carried out with Time (pretest, posttest) and Frequency (high, low) as within-subjects variables and RT on the practice problems as dependent variable. Only a significant effect of Time was found ($F(1, 38) = 29.75$, $p < .001$, $\eta_p^2 = .44$) and no effects of Frequency. This means that the frequency of the practice problems did not influence the results of the children. The effect of frequency was also evaluated in the commutative problems. Of course the commutative problems had not been practiced in a high or a low frequency; therefore, if '5 + 2' was a high frequency practice item, '2 + 5' was called a high frequency commutative item. Again, only an effect of Time was found ($F(1, 38) = 28.16$, $p < .001$, $\eta_p^2 = .43$) and no effects of Frequency.

In order to test whether learning effects were different for children with below average and children with above average mathematical skills, we divided the participants into two groups according to the 50th percentile of their mean reaction time on the pretest. Nineteen children with slow reaction times (M = 3882 ms, SD = 821) were labeled as having 'poor' mathematical skills and twenty children with fast reaction times (M = 2091 ms, SD = 413) were labeled as having 'good' mathematical skills. A repeated measures Anova was carried out with Time (pretest, posttest) and Problem (practice, commutative, related, unrelated) as

[1] Problem '2 + 3' was not included in the scores, because in both versions of the training booklet this problem was regular in frequency.

within-subjects variables, and Level (poor skills, good skills) as between-subjects variable. There was of course a main effect of Level, because we labeled the children according to their performance. Furthermore, there was a significant interaction effect between Level and Time ($F(1, 37) = 4.54$, $p < .05$, $\eta_p^2 = .11$): children with poor skills improved more from pretest to posttest (mean difference 566 ms) than children with good mathematical skills (mean difference 265 ms). Most important, however, there was no effect of Level × Problem × Time. This finding suggests that the pattern of learning effects was similar for children with poor and children with good mathematical skills.

In summary, the results clearly show that the practice participants received was effective. The children also improved on the commutative problems. This improvement was slightly, but not significantly, less than the improvement on the practice problems. These findings suggest a transfer of the knowledge gained during practice to the unpracticed commutative problems. In contrast, there was no transfer of knowledge to the related problems. Although the children improved on the related problems, this improvement was much lower compared to the first two categories. Furthermore, the improvement on the related problems was not significantly different from the improvement on the unrelated control problems. The frequency of the problems in the training booklets did not seem to influence the learning effects. With respect to differences in mathematical ability, children with below average mathematical skills showed the same pattern of learning effects as children with above average mathematical skills.

Experiment 2 is a replication of Experiment 1, but with multiplication problems instead of addition problems. The participants are children in Grade 3 who recently started working with multiplication problems. In this experiment, participants will be tested not only after the second week of practice, but also after the first week of practice. After the first week, children will then have less experience with low frequency problems, so that results for low and high frequency problems may be different. Furthermore, we will make some changes to the set up of the tests; there is a time limit for answering and the experimenter will record whether children answered in time. The time limit is introduced because the problems are relatively novel to the participants. Certain problems may be too difficult for the children to solve and we do not want them to plod on one problem for a very long time. Because of the different answer procedure, the dependent variable will be accuracy within the time limit instead of reaction time.

5.3 Experiment 2: Multiplication

5.3.1 Method

5.3.1.1 Participants

The participants were 67 Grade 3 students (31 girls, 36 boys) from two Dutch schools for primary education. Mean age of the children was 8.6 years (*SD* =

0.48). The children were tested in the beginning of the school year. In the Netherlands, the mathematics curriculum is roughly the same on every school. Generally, at the time of this study the children were supposed to be familiar with the multiplication tables of 1, 2, 3, 4, 5 and 10, but not yet with the tables of 6, 7, 8, and 9. However, on one of the schools it was decided to teach the multiplication tables from 6 to 9 a few months earlier, so the children from school 1 (21 children) had more experience with the test problems than the children from school 2 (46 children). The Dutch mathematics curriculum emphasizes the relationships between multiplication facts. If a child cannot retrieve the answer of a multiplication problem, it is encouraged to use known facts (e.g., $9 \times 8 = [10 \times 8] - 8$ or $4 \times 6 = [2 \times 6] \times 2$) to solve the problem.

Scores on a standard mathematics test (Cito, 2005) showed that the participants had relatively high skills in mathematics. This test is part of a student monitoring system and was administered at the end of the previous school year (about 3 months prior to the pretest). The scores were obtained from the schools; 3 children did not have a result for the test because they were not present at the time of testing. Because only 19 children had scores beneath the 50th percentile of the normalized scores, dividing the children into a group of ‘poor’ and ‘good’ students relative to 50th percentile would result into two groups very unequal in number. Therefore, for the purpose of analyzing the data of the study, we divided the children into three groups with poor, average, and good mathematical skills. Children with scores beneath the 50th percentile of the normalized scores were classified as ‘poor’, children between the 50th and 75th percentile were classified as ‘average’ and children above the 75th percentile were classified as ‘good’. In total, 19 children were classified as poor, 20 as average, and 25 as good.

5.3.1.2 Material and procedure

The participants practiced for two weeks on a set of 6 multiplication problems from the multiplication tables up to 10. Children were tested on their performance before the practice period, after one week of practice, and again after two weeks of practice. Practiced and unpracticed problems are listed in Table 5.2. Again, the unpracticed problems came from three categories: 1) commutative problems, 2) related problems, and 3) unrelated problems. In the commutative problems the first and second factors of the practice problems were switched. In the related problems the first factor of the practice problems was increased with one unit. The unrelated problems were unrelated to the practice problems in terms of the previous two categories.

Table 5.2. Multiplication problems presented on the pretest and the two posttests. During the practice period, children practiced only the practice problems. Items '3 × 8' and '8 × 7' were high in frequency in version A of the training booklet; items '7 × 4' and '6 × 9' were high in frequency in version B.

Practice problems	Commutative problems	Related problems	Unrelated problems
6 × 3 = 18	3 × 6 = 18	7 × 3 = 21	3 × 7 = 21
3 × 8 = 24	8 × 3 = 24	4 × 8 = 32	4 × 6 = 24
7 × 4 = 28	4 × 7 = 28	8 × 4 = 32	9 × 3 = 27
7 × 6 = 42	6 × 7 = 42	8 × 6 = 48	9 × 4 = 36
6 × 9 = 54	9 × 6 = 54	7 × 9 = 63	6 × 8 = 48
8 × 7 = 56	7 × 8 = 56	9 × 7 = 63	8 × 9 = 72

A similar procedure was used as in Experiment 1, except for two changes. The experimenter pressed the response button as soon as soon as the child gave an answer and there was a response time limit of 8 seconds. This time limit is expected to encourage the use of a retrieval strategy or a quick procedural strategy, because there was not sufficient time to solve, for instance, 8 × 7 by repeated addition. After the problem had been presented on the screen for 4 seconds, four horizontal bars appeared at the bottom of the screen. Every second the rightmost bar 'ticked away' to indicate the time left to answer. Children were encouraged to answer within the time limit. If an answer was outside the time limit, it was labeled incorrect. The score of the children consisted of the number of correct answers within the time limit. On average, children needed about 10.5 minutes to finish the pretest and about 8.5 minutes to finish each of the posttests.

Training material were A4 size booklets with 1200 multiplication problems, comprised of 120 blocks with 10 problems. Two different versions of the training booklets were randomly assigned to the children. In version A problems '3 × 8' and '8 × 7' were listed twice as frequent as the other four problems and in version B '7 × 4' and '6 × 9' were listed in a double frequency. A block of 10 problems consisted of 8 practice problems (6 items + 2 items repeated) and 2 filler items. Filler items were relatively easy multiplication problems with 1, 2, 5, or 10 as one of the factors. The maximum number of problems in one session was 200. Furthermore, on the last page of the booklets a list of the multiplication tables from 1 to 10 was printed. This 'crib sheet' was introduced to prevent children from making too many errors, because the selected practice problems were relatively novel to them. If a child did not know the answer to a problem, he/she could look the answer up on the last page of the booklet. The instruction emphasized the importance of making few errors and that, when in doubt, children should look the answers up. For two weeks, children practiced three times a week with their own teacher; each session lasted for 10 minutes. Prior to session 2, children studied the mistakes they made in session 1 and tried to correct them.

5.3.1.3 Analysis

Because children from school 1 had more experience with multiplication problems than children from school 2, children from school 1 had significantly higher scores on the pretest, $F(1, 65) = 6.01$, $p < .05$, $\eta_p^2 = .09$. On average, children from school 1 answered 28 (out of 48) items correctly and children from school 2 answered 20 items correctly. For this reason, we will include ‘school’ as a between-subjects factor in the analyses to control for possible interaction effects. During the practice period, children made 627 problems on average (including fillers). The mean percentage of errors during practice was 6.5 %. Performance improved during the practice period; in the first session the mean percentage of errors was 10.1 % and in the last session it was 4.8 %. There were no significant differences between the schools in how many problems the children made or in the percentage of errors during practice.

5.3.2 Results

A repeated measures Anova was carried out with Time (pretest, posttest 1, posttest 2) and Problem (practice, commutative, related, unrelated) as within-subjects variables. The dependent variable was the number of correct items within the prefixed time limit of 8 seconds. Initially, School and Version (training booklet A or B) were included as between-subjects variables. The results showed no main effect for School or Version and no significant interaction effects with Version. There was one significant interaction effect with School: Problem × School ($p < .01$, with a Greenhouse-Geisser correction). This means that children from school 1 scored differently on the four problem categories than children from school 2. However, because there was no significant interaction effect between Time, Problem, and Version, we decided to leave both School and Version out of the final analysis.

The results of the pretest and the two posttests are presented in Figure 5.2. The maximum score was 12, because the 6 problems in each category were presented twice. The repeated measures Anova showed a main effect of Time ($F(2, 132) = 104.93$, $p < .001$, $\eta_p^2 = .61$): mean scores increased with each test (5.5 on the pretest, 7.4 on posttest 1, and 8.2 on posttest 2). There was also a main effect of Problem ($F(3, 198) = 21.17$, $p < .001$, $\eta_p^2 = .24$) and a significant interaction between Time and Problem ($F(6, 396) = 18.17$, $p < .001$, $\eta_p^2 = .22$). All effects were corrected for sphericity with a Greenhouse-Geisser correction. No significant differences between the four problem categories were present on the pretest. Within-subjects contrasts showed that the improvement for commutative problems was not significantly different from the improvement for practice problems; in both categories children scored about 3 items higher. Furthermore, the improvement for related problems (0.6 items more) did not differ from unrelated problems (0.9 items more). In contrast, the improvement for practice problems was significantly different from related and unrelated problems and also the improvement for commutative problems was significantly different from related and unrelated problems (all $p < .001$). As can be seen in Figure 5.2, children learned most in the first week of practice, although in the second week scores for all four problem categories still improved a little. There was a significant effect of Time between posttest 1 and posttest 2 ($F(1, 66) = 34.01$, $p < .001$, $\eta_p^2 = .34$), indicating that scores improved. However, there was no significant interaction between Time and Problem, which means that improvement for all problem categories was the same from posttest 1 to posttest 2. Compared to the pretest, scores for all four problems categories improved on posttest 2 ($p < .001$). Although on posttest 2 the score for practice problems was 0.5 item higher than the score for commutative problems, within-subjects contrasts showed that the improvement for commutative problems from pretest to posttest 2 did not significantly differ from the improvement for practice problems. In fact, all within-subjects contrasts were statistically the same as measured on posttest 1.

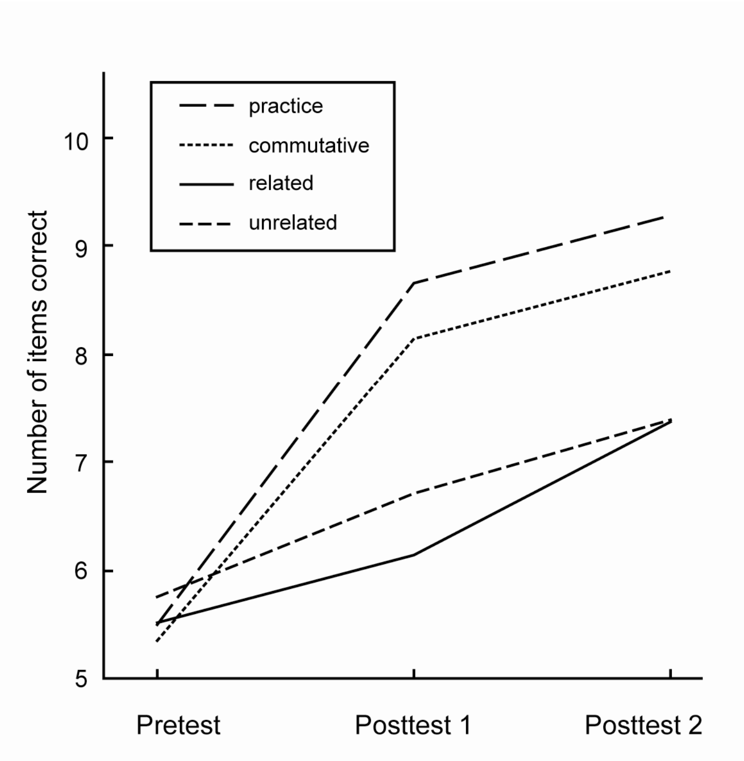

Figure 5.2. Multiplication problems in Grade 3. Mean scores (number of items correct within 8 seconds) on the pretest and the two posttests for practice, commutative, related, and unrelated problems. Maximum score is 12.

To evaluate frequency effects, separate scores were calculated for practice problems with a double frequency and practice problems with a regular frequency[2]. A repeated measures Anova was carried out with Time (pretest, posttest 1, posttest 2) and Frequency (high, low) as within-subjects variables. In

[2] Problems '6 × 3' and '7 × 6' were not included in the scores, because in both versions of the training booklet these two problems were regular in frequency.

the first analysis the score on the practice problems was the dependent variable and in the second analysis the score on the commutative problems. In both analyses, only an effect of Time was found ($p < .001$). This means that there were no effects of frequency.

Finally, we tested whether children with poor, average, and good mathematical skills showed different learning effects. A repeated measures Anova was carried out with Time (pretest, posttest) and Problem (practice, commutative, related, unrelated) as within-subjects variables, and Level (poor skills, average skills, good skills) as between-subjects variable. There was a main effect of Level ($p < .05$): children with better skills had higher scores. Furthermore, there was a significant interaction effect between Level and Problem ($p < .05$), indicating that children with different skills scored differently on the four problem categories. However, there was no effect of Level × Problem × Time. This finding suggests that the pattern of learning effects was similar for children with different abilities in mathematics.

In sum, the results of the multiplication experiment show that the practice participants received was effective. The children also improved on the commutative problems; this improvement was slightly less than the improvement on the practice problems, but the analysis of variance indicated no significant difference between the development of scores for practice and commutative problems. These findings suggest a transfer of the knowledge gained during practice to the unpracticed commutative problems. Children also improved on the other unpracticed problems – related and unrelated problems – but this improvement was much lower. Furthermore, the improvement on related problems was not significantly different from the improvement on the unrelated control problems. Results were similar for children with different abilities in mathematics. Just as in Experiment 1, there were no effects of frequency. The largest learning effects occurred after the first week of practice.

5.4 General Discussion

In two experiments transfer effects in simple addition and multiplication were studied in primary school children. Children practiced for two weeks on a small set of problems. Before and after practice, they were tested on four problem categories: 1) practice problems, identical to the problems presented during practice, 2) commutative problems with an operand order change (e.g., 6 + 3 practiced, 3 + 6 on the test; 7 × 4 practiced, 4 × 7 on the test), 3) related problems with one of the operands increased with one unit (e.g., 6 + 3 practiced, 6 + 4 on the test; 7 × 4 practiced, 8 × 4 on the test), and 4) unrelated problems. The unrelated problems served as a control for general speed up, for instance because children were more familiar with the testing procedure on the posttest, or because they developed faster procedural skills during the practice period. The results were similar for multiplication and addition. The commutative problems improved almost as much as the practice problems; no significant differences were found between the two categories. This indicates a positive transfer of the knowledge gained during the practice period to the commutative problems. In contrast, there was no transfer effect for related problems; children improved hardly on this category and improvement was not higher than improvement for unrelated problems. Children with different abilities in mathematics showed similar learning effects.

The results suggest that the Identical Elements (IE) model for arithmetic fact representation (Campbell et al. 2006; Rickard, 2005; Rickard & Bourne, 1996; Rickard et al., 1994) applies not only to adults, but also to children who are starting to master simple arithmetic. Adults have more mental resources and have strategies available for arithmetic problem solving that are far more efficient than those of children. Consequently, it was not immediately evident that transfer effects from practice would be similar in adults and children. According to the IE model, there is a single long-term memory node for problems consisting of the same numerical elements (i.e., operands and answer), regardless of operand order. When the numerical elements of a test problem do not match exactly with those of a practice problem, there will be no positive transfer. In the present study positive transfer was found when children were tested on problems with the same numerical elements (practice and commutative problems), but not when the numerical elements were different (related and unrelated problems). The findings suggest that children have a single representation in long-term memory for addition and multiplication problems that differ only in operand order.

Furthermore, because there was no transfer effect from practice problems to related problems, it seems that participants did not use a derived fact strategy for related problems (e.g., solving 6 + 4 as [6 + 3 = 9] + 1 when 6 + 3 was practiced). Also, the results were similar for children with different abilities in mathematics, so even children with good mathematical skills were not able to benefit from the knowledge gained in the practice period when asked to solve related problems. Although the related and unrelated problems were slightly larger in problem size than the practice and the commutative problems and there unevitably were differences in the position of the larger operand, it is not likely that this influenced the learning effects of the children. The main finding to conclude that no contamination of possible confounds occurred is that there was no significant difference between the four problem categories on the pretest of both the addition and the multiplication experiment.

In the present study, improvement on the practice problems was slightly higher than improvement on the commutative problems, but this difference was not significant. According to Rickard et al. (1994), the IE model is not necessarily inconsistent with a small difference in performance between practice and commutative problems, because the model assumes distinct perceptual and cognitive stages. In the cognitive stage, there is an abstract representation that does not depend on operand order. However, in the perceptual stage that precedes the cognitive stage, there may be an advantage for problems with the same operand order as practiced. An alternative explanation to account for the small difference between practice and commutative problems is that children have both versions of the problem available in long-term memory and they rapidly transformed the operand order in the case of the commutative problems. Based on the current data we cannot distinguish between the two explanations. However, because the difference in improvement between practice and commutative problems was small and not significant, we consider it plausible that children have a single representation for both operand orders.

The most straightforward educational implication from the present study is that if teachers want children to practice simple additions and multiplications, it is not necessary to let them practice both commutative orders. The results show clearly that if children practice simple arithmetic problems with the operands in one specific order, they will also improve on unpracticed problems with the operands in the opposite order. This reduces the number of simple arithmetic problems that need to be practiced by half. There are also some important theoretical implications. The first implication is that the IE model is not only useful to predict performance when the associative network for arithmetic facts is more or less established in long-term memory, but also when arithmetic facts are relatively new, because the children in the present study were still in the process of memorizing simple arithmetic facts. The second theoretical implication is that the storage of arithmetic facts in memory starts at an early age. In accordance with previous research (Lemaire et al., 1994; Siegler, 1988), the results suggest a memory-based process; children did not simply become faster in procedural strategies. The third theoretical implication is that it is possible to model the acquisition of simple arithmetic facts in a relatively short learning experiment, which seems promising for further research.

Contrary to our expectations, no frequency effects were found. The fact that some problems were practiced with a double frequency did not seem to influence improvement on those problems. This was an unexpected finding, because it is often assumed that frequency effects shape the associative network for multiplication facts (Ashcraft & Christy, 1995; Geary, 1996; Siegler, 1988). One explanation for the absence of frequency effects in the current study is that even the regular frequency problems were practiced many times and that this caused a ceiling effect. To investigate this hypothesis, an extra posttest after one week of practicing was included in the multiplication experiment. The results from the two posttests showed that children learned most in the first week of practice and that differences between problem categories were already apparent then. Furthermore, the learning and transfer effects that were found after the first week of practice did not change in the second week. Finally, frequency effects were absent, also after the first week of practice. It is of course possible that there were not sufficient items (2 problems with a regular frequency and 2 problems with a double frequency, presented two times on the tests) to measure frequency effects reliably. However, it is also still possible that a ratio of 1:2 is too small a difference to induce frequency effects.

The present study with primary school children shows that mechanisms involved in storing addition and multiplication facts are highly similar. Similar effects in retrieving simple addition and multiplication facts have earlier been found in adults' performance (see Campbell, 1995, for an overview). For instance, both addition and multiplication problems suffer from the problem size effect: people take longer and make more errors if the digits in an arithmetic problem (operands and answer) are larger in numerical size. Also, tie problems with a repeated operand are easier to remember than non-tie problems, both in multiplication (e.g., 3×3) and in addition (e.g., $4 + 4$). Although there is evidence that addition and multiplication facts are stored in separate semantic memory networks in adults (Van Harskamp & Cipolotti, 2001), the finding that some production errors stem from choosing the wrong operation (e.g., $6 \times 7 = 13$; Miller, Perlmutter, & Keating, 1984) suggests that these networks are interrelated. The present study shows that fact retrieval in addition is similar to multiplication not only in adults, but also in children. It is interesting that results were so similar for addition and multiplication, because instruction in the Netherlands emphasizes the use of retrieval much more in the case of multiplication than in the case of addition. However, it must be noted that the experiment on addition was conducted when children were already familiar with solving addition problems up to ten. At that time, participants were probably starting to use retrieval more often, so we assume that the results reflect mainly retrieval processes instead of procedural strategies. Therefore, we can conclude that, just as in adults, retrieval processes in children are highly similar in addition and multiplication.

In summary, the present study tested the predictions of the identical elements model for arithmetic fact representation (Campbell et al. 2006; Rickard, 2005; Rickard & Bourne, 1996; Rickard et al., 1994) in primary school children. A transfer effect was found for problems with an operand order change, but not for problems where one of the operands was increased with one unit. The same results were found for addition and multiplication, indicating that processes for retrieval in both domains are highly similar. The results support the identical elements model, which assumes a single long-term memory node for problems consisting of the same numerical elements, regardless of operand order.

5.5 Acknowledgments

We are grateful to all the students and teachers who participated in this study and to Andrea van der Laan and Jolanda Geerts for their help in collecting the data.

6

Can individual differences between children explain their ability to learn arithmetic facts?

Recent studies show that individual differences in working memory, counting speed, and rapid naming are related to mathematical ability. This study used a learning task design to study the relationships between cognitive processing and mathematical achievement. Grade 1 children (7.2 years) performed a learning task with simple addition problems and Grade 3 children (8.7 years) performed a learning task with simple multiplication problems. The results show that most children improved on the practiced problems. It was not possible to predict individual learning effects on the practice task from differences between children on measures of cognitive processing. However, correlational analyses revealed that Digit span forward, Digit span backward, Counting speed, and Rapid naming seem to be related to mathematical ability. Furthermore, a domain-specific relationship was found between verbal short-term memory span and mathematical ability.

6.1 Introduction

Learning to solve basic arithmetic problems accurately and with little effort is an important goal in primary education, because fast retrieval of answers to simple problems (e.g., 3 + 6 = …) is a prerequisite for solving more complex problems (e.g., 13 + 26 = …) (Adams & Hitch, 1997; Cumming & Elkins, 1999). Through extensive practice with addition problems, children in primary school move on from initial counting procedures to faster memory-based strategies (Geary, Bow-Thomas, Liu, & Siegler, 1996; Goldman, Mertz, & Pellegrino, 1989; Siegler, 1987). A similar development occurs in multiplication; with increasing experience, strategies like repeated addition (e.g., 3 × 4 = 4 + 4 + 4) or counting sequences (e.g., 3 × 5 = 5, 10, 15) are replaced by retrieval (Cooney, Swanson, & Ladd, 1988; Mabbott & Bisanz, 2003; Lemaire & Siegler, 1995). Frequent repetition of problems is an important factor in this process (Ashcraft & Christy, 1995; Geary, 1996; Siegler, 1988). Unfortunately, not all children succeed in learning arithmetic facts. Children with mathematical disabilities often have particular difficulty in memorizing arithmetic facts (Andersson, 2008; Geary, 1993; Hanich, Jordan, Kaplan, & Dick, 2001) and these difficulties persist over time (Jordan, Hanich, & Kaplan, 2003; Ostad, 1998). In the present study individual characteristics of children in Grade 1 and Grade 3 were studied that may explain why some children have difficulties learning arithmetic facts and others have not. Correlational analyses were used to study the relation between performance of children on cognitive processing tasks and their performance on a learning task with basic addition or multiplication problems.

6.1.1 Working memory and short-term memory

One of the most consistent findings in the literature on children with mathematical disabilities is that working memory is involved in learning mathematics. To explain why working memory is so important, we must first define what working memory is. The theoretical framework most often used is Baddeley’s multi-component working memory model (Baddeley & Hitch, 1974). This model includes two storage systems, the phonological loop (for verbal information) and the visual-spatial sketchpad (for visual-spatial information). A third component, the central executive, is assumed to control and manage working memory processes. A defining characteristic of the working memory system is that it is limited in its capacity for storage and processing. Furthermore, it is important to distinguish between working memory and short-term memory. Short-term memory involves storage of information, whereas working memory involves both storage and processing. An example of a short-term memory task is forward digit span, in which participants are required to listen to a string of digits and then reproduce them in order. Backward digit span, on the other hand, is a working memory task, because participants reproduce the string of digits in the reverse order. Such a task requires participants not only to store information, but also to control and update the information in working memory.

There is strong evidence that mathematical disabilities are related to deficits in working memory (D'Amico & Guarnera, 2005; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; McLean & Hitch, 1999; Passolunghi & Siegel, 2004; Rosselli, Matute, Pinto, & Ardila, 2006; Siegel & Ryan, 1989). Among researchers there is wide consensus that children with mathematical difficulties show impairments on complex working memory tasks that involve the central executive. An example of such a task is Counting span, in which participants are required to count a subset of the items on a display (e.g., all red circles among the blue circles) and to remember the number counted. A series of such displays is shown and the counting span is the longest sequence for which participants can remember all the counts. Although the usage of a task like Counting span implies that the central executive is viewed as a unitary system, more recent research suggests that several executive functions can be distinguished from each other, such as inhibition, shifting, and updating, and that these functions may have different relationships with mathematical ability (Baddeley, 1996; Deschuyteneer & Vandierendonck, 2005; Van der Sluis, De Jong, & Van der Leij, 2004).

It is not entirely clear whether the working memory problems of children with mathematical difficulties are domain-general or domain-specific. Siegel and Ryan (1989) found that children with mathematical difficulties but with normal reading ability scored lower than controls on a Counting span task, but not on a task usually called Listening span, which involves verbal instead of numerical stimuli. The Listening span task requires participants to state whether sentences are true or false and to remember the last words of a series of sentences. The results of Siegel and Ryan (1989) were not replicated by Passolunghi and Siegel (2004), who found an impaired performance on both Counting span and Listening span for children with mathematical difficulties but no reading problems. Nevertheless, in this study a domain-specific deficit was found on two working memory tasks that were more simple than Listening span and Counting span, namely Digit span backward and Word span backward. In the Word span backward task familiar bisyllabic words were used as stimuli instead of digits. Children with mathematical difficulties scored lower than controls on the Digit span backward task, but not on the Word span backward task. Other researchers have argued that both domain-general and domain-specific aspects of working memory are related to mathematical ability (Swanson & Sachse-Lee, 2001; Wilson & Swanson, 2001). Whether a domain-specific or a domain-general deficit in working memory is found will probably depend on the selection of participants, because the group of children with mathematical difficulties is very heterogeneous (Geary, 1993). For instance, results may be influenced by whether a strict or a lenient mathematics achievement cut-off score is used (Geary et al., 2007) or the method that is used to control for differences in reading ability (Bull & Johnston, 1997), because reading and mathematical disabilities often co-occur (Geary, 1993).

Although there is consensus that the central executive is involved in mathematical problem solving, the roles of the phonological loop and the visual-spatial sketchpad are less clear. Most research has focused on the phonological loop. Usually, verbal short-term memory tasks like Digit span forward or Word span forward are used as an index of the phonological loop. In addition, sometimes Nonword repetition is used as a measure of phonological short-term memory. Some studies have found that the phonological loop is impaired in children with mathematical difficulties (Koontz & Berch, 1996; Swanson & Sachse-Lee, 2001; Webster, 1979) whereas other studies found no relation between the phonological loop and mathematical ability (Bull & Johnston, 1997; McLean & Hitch, 1999; Passolunghi & Siegel, 2004). In other words, there is still debate on the role of the phonological loop in mathematics.

With respect to the visual-spatial sketchpad, over the last years increasing evidence was found that performance on visual-spatial short-term memory tasks is related to mathematical achievement (D'Amico & Guarnera, 2005; Holmes & Adams, 2006; McLean & Hitch, 1999). A task regularly used as an index for visual-spatial short-term memory is Corsi blocks span (Corsi, 1972). This task is similar to Digit span forward, but the stimuli consist of blocks that are tapped in sequence. Other tasks for visual-spatial short-term memory require participants to remember a route through a maze or to remember the position of coloured blocks in a visual matrix. Working memory tasks with visual-spatial stimuli, like Mental rotation, are not commonly used, although there is one study (Reuhkala, 2001) that found a relation between mental rotation and mathematical ability in adolescents (ages 15 to 16). In a study with normal primary school children (not screened for mathematical difficulties), Holmes and Adams (2006) found that scores on a central executive task (Listening span) and on a visual-spatial sketchpad task (Mazes memory) predicted unique variance in children's mathematical achievement scores, whereas scores on a phonological loop task (Nonword span forward) did not. The role of the visual-spatial sketchpad was particularly strong in the younger participants (mean age 8.1 years). Interestingly, when the data of the older participants (mean age 9.8 years) were reanalyzed with separate analyses for easy and difficult problems, it was found that the phonological loop did predict some unique variance on the scores for the easy problems. This finding is consistent with the idea that children rely on verbal codes when they use a retrieval strategy for easy arithmetic problems, at least for multiplication facts (Dehaene & Cohen, 1997; Lee & Kang, 2002).

How is the finding of a working memory deficit in children with mathematical difficulties related to the observation that these children have particular problems with learning arithmetic facts? In normal mathematical development, children appear to develop associations between a problem and its solution automatically through the use of nonretrieval strategies, such as counting (Siegler, 1988). Geary, Brown, and Samaranayake (1991) suggested that a working memory deficit may contribute to the failure of children with mathematical difficulties to develop adequate representations of basic arithmetic facts in long-term memory. The original representation of the problem's integers may decay so quickly that the answer fails to become associated with the problem, even after extensive practice. For this reason, children with mathematical difficulties continue to rely on counting-based strategies, even though retrieval strategies are much quicker.

Working memory becomes less important once representations of arithmetic facts have been established in memory (Geary et al., 2004).

6.1.2 Counting speed

Counting speed (the speed of reciting for instance the numbers 1-20) is assumed to be related to verbal short-term memory. The explanation for this relation is that the ability to recite numbers quickly offers an advantage in the Digit span forward task, because it is possible to keep more digits active in the phonological loop. Evidence for this hypothesis was found by Geary, Bow-Thomas, Fan, and Siegler (1993) who showed that Chinese kindergartners (mean age 5.9 years), had a larger forward digit span than American children of the same age. Chinese number words are shorter than English number words and can therefore be counted more quickly. In the same study the Chinese children were found to outperform the American children on simple addition. Geary et al. (1993) attributed this better performance to the higher digit span of the Chinese children, because both Chinese and American children with a higher digit span tended to use verbal counting instead of slower finger counting strategies.

It is unclear whether children with mathematical difficulties are slower in counting. Hitch and McAuley (1991) found a slower counting speed in children with mathematical difficulties, but Passolunghi and Siegel (2004) did not. However, counting speed may well be related to arithmetic fact learning. The use of slow counting procedures could result in the phonological representations of the addends decaying in short-term memory before the count is completed, and therefore reducing the likelihood that the problem and the answer become associated in long-term memory.

6.1.3 Rapid automatized naming (RAN)

Rapid automatized naming (RAN) is usually tested with a visual array of 50 stimuli, consisting of five symbols in a given category (e.g., letters, numbers, colours, objects) that are presented 10 times in random order (Denckla & Rudel, 1974). Participants are required to name the symbols as fast and accurate as possible. In other words, RAN involves the rapid naming of highly repetitive stimuli. The speed of RAN provides an index of the ease of retrieving phonological representations from long-term memory. There is extensive evidence that reading speed is related to RAN, especially to letter and digit naming (Van den Bos, Zijlstra, & Spelberg, 2002; Wolf & Bowers, 1999). However, much less is known on the relation between RAN and mathematical problem solving. Van der Sluis et al. (2004) found that children with reading difficulties were slower than children without reading difficulties in the naming of letters and digits. Children with mathematical difficulties, on the other hand, were slower in the naming of digits only compared to children without mathematical difficulties. Furthermore, they were also slower on a rapid naming task that required participants to decide on the quantity of groups of objects. Similar results were found by Bull and Johnston (1997). They found no difference in the speed of identifying letters between a group with low mathematical ability and a group with high mathematical ability, once reading ability was controlled for. In contrast, the speed of identifying digits was related to mathematical ability.

There are at least two possible reasons why slower digit naming may be related to less efficient arithmetic fact learning. In the first place, slow access to phonological representations of numbers in long-term memory may indicate an impairment in basic numerical processing, which potentially could lead to all sorts of mathematical problems. Secondly, slow digit naming could be a symptom of a more general problem in the storage and access of numerical material in long-term memory, including basic arithmetic facts.

6.1.4 The present study

In the present study a learning task design is used to study the relationship between cognitive processing and the ability to learn basic arithmetic facts. A learning task measures what a child is able to learn in a short period of time. By using a learning task with repeated practice on simple arithmetic problems we hope to simulate the development of arithmetic fact learning over a longer period of time. Instead of studying children with mathematical disabilities, as is usually the case in the literature, our participants were children with average mathematical abilities, because much less is known about the role of cognitive variables in normal mathematical development. Children in Grade 1 (mean age 7.2 years) performed a learning task with simple addition problems and children in Grade 3 (mean age 8.7 years) performed a learning task with simple multiplication problems. Apart from the learning task, the children were tested on cognitive processing tasks that involved short-term memory, working memory, counting speed, and rapid automatized naming. Furthermore, they were tested on one or two tasks that involved the learning of arbitrary associations. The reason for including these tasks is that for learning arithmetic facts children need to form associations between problems and answers (e.g., Siegler, 1988). If children who are poor at learning arithmetic facts also have difficulties with forming associations in a task with non-numerical material, there may be a more general problem. Correlational analyses were used to study the relationship between individual differences on the cognitive processing tasks and children’s performance on the learning task. Furthermore, we studied the relations between individual scores for cognitive processing and general scores for mathematical achievement.

6.2 Experiment 1: Addition

6.2.1 Method

6.2.1.1 Participants

The participants were 38 Grade 1 students (13 girls, 25 boys) from five Dutch schools for primary education. Mean age of the children was 7.2 years (*SD* = 0.45). Participants were selected with a mathematics test that is commonly used in the Netherlands (De Vos, 1992). Children performed 1 minute of additions and 1 minute of subtractions, with problems increasing in difficulty. The score was the total number of problems answered correctly. Selected children had below-average to average scores, relative to the mean score in their school class. However, children with really low scores (less than 10 problems correct in two minutes) were not selected, because participants had to be able to solve the addition problems in the practice task without help. Participants' mean score on the mathematics test was 15 problems (*SD* = 2.9) answered correctly in two minutes. As a second measure of mathematical ability the score on the most recent administered mathematics test of a pupil monitoring system (Cito, 2002) was obtained from the schools. The pupil monitoring system tests a broader range of arithmetic skills: it includes subjects like counting, ordering numbers, and breaking up a number into smaller numbers. Most of the problems are provided within a realistic context. The tests are administered twice in every school year, according to an established schedule in the schools. The most recent test scores were obtained 2-3 months before the experiment was carried out. The normalized scores showed that 18 participants scored in the lowest quartile, 8 in the second quartile, 9 in the third quartile, and 3 in the highest quartile of the normal population.

6.2.1.2 Academic achievement measures

Four measures of academic achievement were included as predictors for a learning effect in the practice task.

1. Mathematics general. This was the score from the mathematics test of the pupil monitoring system (Cito, 2002) and was obtained from the schools.

2-3. Addition / Subtraction. The score consisted of the number of addition problems answered correctly in one minute and the number of subtraction problems answered correctly in one minute. The mathematics test (De Vos, 1992) was administered to a whole school class at the same time.

4. Reading skill. Reading skill was assessed with a commonly used Dutch reading test (Brus & Voeten, 1973). The test requires the participant to read as many words as possible within one minute, with words increasing in difficulty. The score consisted of the number of words read correctly in one minute.

6.2.1.3 Cognitive processing measures

Seven measures of cognitive processing were included as predictors for a learning effect in the practice task.

1. Counting speed. Children were instructed to count forward from 6 to 13 and from 11 to 18, and backward from 10 to 3. Responses were timed on a stopwatch from the first to the last item, measured to the nearest tenth of a second. The three counting scores were added together to obtain one score for counting speed. Intercorrelations between the three counting measures were: r = .71 (p < .001) between 6-13 and 11-18, r = .31 (p < .05) between 6-13 and 10-3, and r = .61 (p < .001) between 11-18 and 10-3.

2-3. Digit span forward / backward. Digit span forward and digit span backward were assessed using subtests from WISC-III (Wechsler, 2002). The experimenter read digit sequences aloud at a rate of approximately one digit per second. In the digit span forward task, participants were required to repeat the digit sequence in the same order as presented. In the digit span backward task, participants repeated the digit sequence in the reverse order (i.e., 5, 7, 4 became 4, 7, 5). The length of the digit sequence was increased with 1 after two trials of the same length. The task started with a span length of two digits and continued until the child made a mistake in both trials of the same span length. Depending on the number of correct trials, the child was rewarded with 2, 1 or 0 points for a certain span length. The score for the task consisted of the total number of digit sequences recalled correctly. Digit span forward has a test-retest reliability of .81 and Digit span backward has a test-retest reliability of .62 (Gathercole, Pickering, Ambridge, & Wearing, 2004).

4-6. Rapid automatized naming (RAN): letters / digits / quantity. Three RAN tasks were administered to assess the child's speed at retrieving phonological representations from long-term memory. The child was asked to name letters, digits, and (following Van der Sluis et al., 2004) decide on the quantity of groups of triangles. Each task had 50 items and comprised of 5 different stimuli: the letter task had letters a, o, p, s, t; the digit task had digits 2, 4, 6, 7, 9; and the quantity task used horizontal arrays of triangles varying in number from 1 to 5 (e.g., 2: ▲▲). The 50 items were printed in five horizontal rows of 10 items each, with spaces in between. The child was required to read aloud the randomly ordered items as accurately and quickly as possible. A stopwatch was used to measure the child's naming time to the nearest tenth of a second. The task ended when all items had been named or when one minute had passed. The score consisted of the average time needed per item (in seconds). The average test-retest reliability of RAN letters and digits is .74 (Van den Bos et al., 2002).

7. Learning names. Children's memory for learning associations was assessed using a subtest of the RAKIT (Bleichrodt, Drenth, Zaal, & Resing, 1984), a Dutch intelligence test for children. The experimenter presented the pictures of 6 butterflies and 6 cats one by one and asked the child to remember the names of the animals. The names were easy to remember in some cases (e.g., “snow white” for a white cat) and arbitrary in other cases. In round 1 the pictures were presented for 10 seconds and the experimenter provided the names of the animals twice. In round 2 the pictures were presented again and the child was asked whether it remembered the name of the animal. If not, the experimenter told the name again. In round 3 the pictures were presented for the last time. The score for this task consisted of the number of correct names in round 2 and 3, i.e., if the child correctly recalled the name of a certain animal in both rounds, it was rewarded with 2 points.

Table 6.1. Addition problems presented on the pretest and posttest. During the practice sessions, half of the children practiced the problems of set A and the other half practiced the problems of set B. Control problems were not practiced.

Set A	Set B	Control
1 + 3 = 4	2 + 3 = 5	1 + 2 = 3
2 + 4 = 6	1 + 5 = 6	2 + 1 = 3
4 + 3 = 7	3 + 4 = 7	3 + 2 = 5
2 + 6 = 8	5 + 3 = 8	4 + 2 = 6
5 + 4 = 9	2 + 7 = 9	2 + 5 = 7
3 + 7 = 10	4 + 6 = 10	5 + 2 = 7

6.2.1.4 Material and procedure of the practice task

Practice sessions. Two practice sets were constructed, consisting of 6 addition problems with answers up to 10. The problems of set A were comparable in problem size to the problems of set B; see Table 6.1 for a list of the two sets. Half of the children practiced set A and the other half practiced set B. The problems were presented on a laptop. A trial started with an empty screen (1500 ms), then an asterisk appeared as a fixation point in the middle of the screen (500 ms), then the screen was empty again (500 ms), and finally the addition problem was presented in the middle of the screen. There was no time limit for answering, but the instruction emphasized both accuracy and speed. As soon as the child answered, the experimenter registered reaction time and accuracy by pressing the left mouse button when the answer was correct and the right mouse button when the answer was incorrect. If a child was overtly counting, the experimenter encouraged the child to solve the problem without counting. After the child had answered, a feedback screen was presented for 1500 ms. The feedback screen was the same for correct and incorrect answers; in both cases the correct answer appeared next to the addition problem on the screen. If the answer of the child was correct, the experimenter regularly said “Very good!” or similar expressions when the feedback screen was presented. If the answer was incorrect, the experimenter encouraged the child to look at the correct answer and remember it for next time. During the practice task, children practiced each problem 12 times.

The material consisted of 12 blocks with 6 practice problems and 2 filler problems. Filler problems were mostly easy addition problems (tie problems and problems with 1 or 2 as one of the addends); see Table 6.2 for the filler problems. The order of the practice problems was pseudo-random: in consecutive trials there was never the same number on the same position in the problem (first addend, second addend, outcome). The problems were practiced in two sessions on two successive days. A practice session started with three filler problems, after which 6 blocks with 48 experimental problems in total were presented. There was a break after every 16 items, during which a photo of an animal was presented. On average, children needed about 8.5 minutes to finish one practice session.

Table 6.2. Filler problems presented during practice sessions.

Filler problems	
3 + 3 = 6	7 + 1 = 8
5 + 1 = 6	6 + 3 = 9
1 + 6 = 7	7 + 2 = 9
6 + 1 = 7	8 + 2 = 10
4 + 4 = 8	9 + 1 = 10

Pretest and posttest. Before and after the practice sessions participants were tested on the problems of set A, set B, and 6 control problems. The control problems were addition problems that could easily be solved by counting; see Table 6.1 for the control problems. Therefore, if participants improved on the practiced problems because of faster counting skills, there should be an improvement as well on the control problems. The test problems were presented in the same way as the practice problems, but during the tests no feedback was provided. The order of the problems was pseudo-random: in consecutive problems there was never the same number on the same position in the problem (first addend, second addend, outcome). All test problems were presented twice. The test started with 3 filler problems, after which 36 experimental trials were presented. Pretest and posttest were identical except for the order of the items. Children needed about 6 minutes to finish a test.

6.2.1.5 General procedure

Each child was tested individually in two sessions of about 20 minutes, conducted on two successive days. Testing took place in a quiet room at school. During the individual test sessions, children performed all cognitive processing tasks and 1 academic achievement task (reading skill). Furthermore, they performed the pretest, two practice sessions, and the posttest of the practice task. All children performed the tasks in the same order; see Table 6.3 for the general procedure.

Table 6.3. Tasks presented to the children in the two individual test sessions.

Tasks day 1	Tasks day 2
1) Counting speed	1) Practice session 2
2) RAN	2) Reading Speed
3) Digit span	3) Learning Names
4) Pretest practice task	4) Posttest practice task
5) Practice session 1	

6.2.1.6 Analysis of the practice task

Practice data. From the practice data, filler trials and trials with an incorrect answer were removed first. Then, for each participant the mean RT and standard deviation were calculated, based on all correct RTs during practice. An RT was categorized as an outlier when it deviated more than 2 standard deviations from the mean RT of the participant. In total, the participants had 11.7 % invalid RTs: 7.5 % incorrect answers and 4.2 % outliers. Outliers were also removed. After this, the mean RT for each of the 6 practice blocks was calculated (the practice blocks were separated from each other with a break; in each practice block the practice problems were presented twice). The mean RT per block was used for the repeated measures analysis of the practice data, see Results. In addition, a measure for improvement during practice was calculated. In a first step, a linear regression analysis was carried out for every participant. The linear regression coefficient can be considered as a measure for improvement during practice, because a steep, negative regression line indicates a fast decrease in reaction time. The regression coefficient correlated negatively with the mean RT during pretest (r = -.59, p < .01), thus indicating that slower children had a larger improvement during practice. This is essentially a scaling problem: slow children seem to improve more because they improve more in absolute terms. We transformed the absolute measure for improvement into a relative measure by dividing the regression coefficient by the mean RT on the pretest. Furthermore, to make the measure for improvement easier to understand, we made negative effects positive and vice versa. This way, a larger positive effect indicated a larger improvement during practice. In short, the measure for improvement during practice was: *− (regression coefficient practice blocks) / mean RT pretest*.

Test data. Originally, 41 children participated. Three children were excluded from the data set because they had too many invalid reaction times during pretest and posttest – they gave an incorrect answer or the RT was categorized as an outlier. The test data was analyzed as follows. Trials with an incorrect answer were removed first. Then, for each participant the mean RT and standard deviation were calculated, based on all correct RTs from the pretest and posttest. An RT was categorized as an outlier when it deviated more than 2 standard deviations from the mean RT of the participant. Outliers were removed. After this, the mean RT was calculated for every problem in the test. Because the problems were presented twice, the mean RT for a problem was based on two RTs. If there was only one valid RT, this RT was taken. If both RTs were invalid, the problem had a missing value. The next step was that if a problem had a missing value on one of the tests, the RT for that problem was also removed from the other test. Children with more than 2 missing values out of the 6 problems of one category (practiced, unpracticed, control) were excluded from the data set – these are the three children that were mentioned before. The remaining 38 children had 11.7 % invalid RTs (7.6 % incorrect answers and 4.1 % outliers), resulting in 6.7 % problems with a missing value. Finally, the mean RT was calculated for each of the three categories in the pretest and posttest. The RT per category was used for the repeated measures analysis of the test data, see Results. In addition, a measure for improvement on the tests was calculated. In a first step the improvement on the practiced problems was calculated: the mean RT for practiced problems on the posttest was subtracted from the mean RT for practiced problems on the pretest. The calculated measure correlated positively with the mean RT on the pretest (r = .34, p < .05), indicating again that slower children had a larger improvement on the tests. Therefore, the measure for improvement was divided by the mean RT on the pretest. In short, the measure for improvement on the tests was *(mean RT practice problems pretest - mean RT practice problems posttest) / mean RT pretest*.

6.2.2 Results

6.2.2.1 Practice task

Practice data. Figure 6.1 shows the mean reaction time for a practice problem in the six practice blocks. In each practice block all practice problems were presented twice. To evaluate whether practice problems were answered faster during practice, a repeated measures Anova was carried out with Block (1 to 6) as within-subjects variable and mean RT as dependent variable. There was a main effect of Block (F(5, 185) = 13.90, p < .001, η_p^2 = .27), indicating that the improvement was significant. Polynomial within-subjects contrasts showed a linear trend in the data of the practice blocks (F(1, 37) = 32.63, p < .001, η_p^2 = .47). There was no significant effect for a quadratic or a cubic trend.

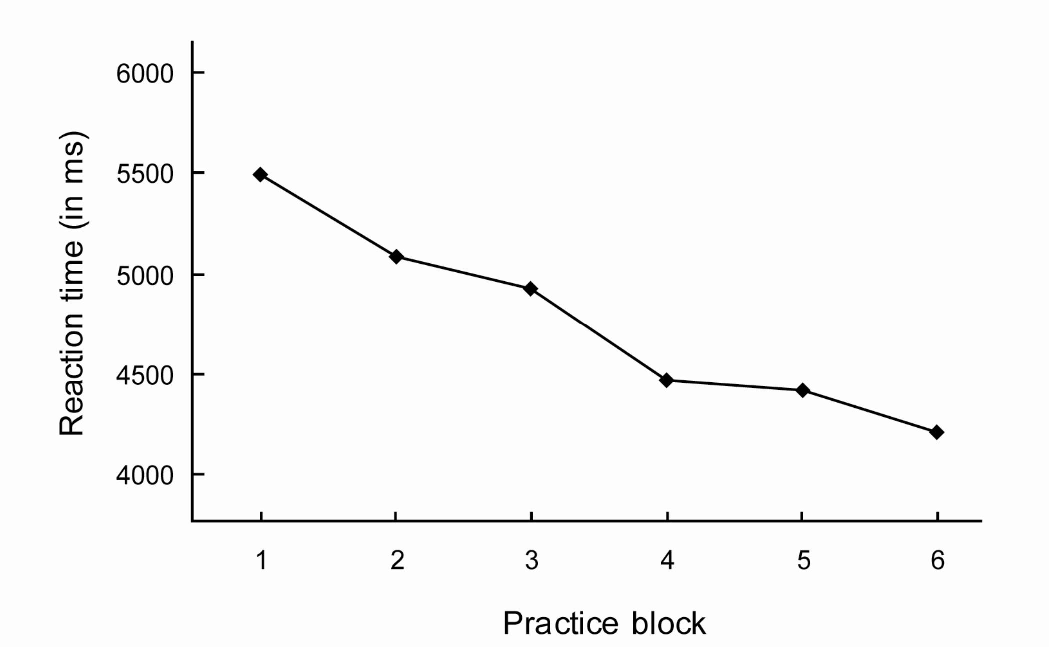

Figure 6.1. Addition problems in Grade 1. Mean reaction times (in ms) in the six practice blocks.

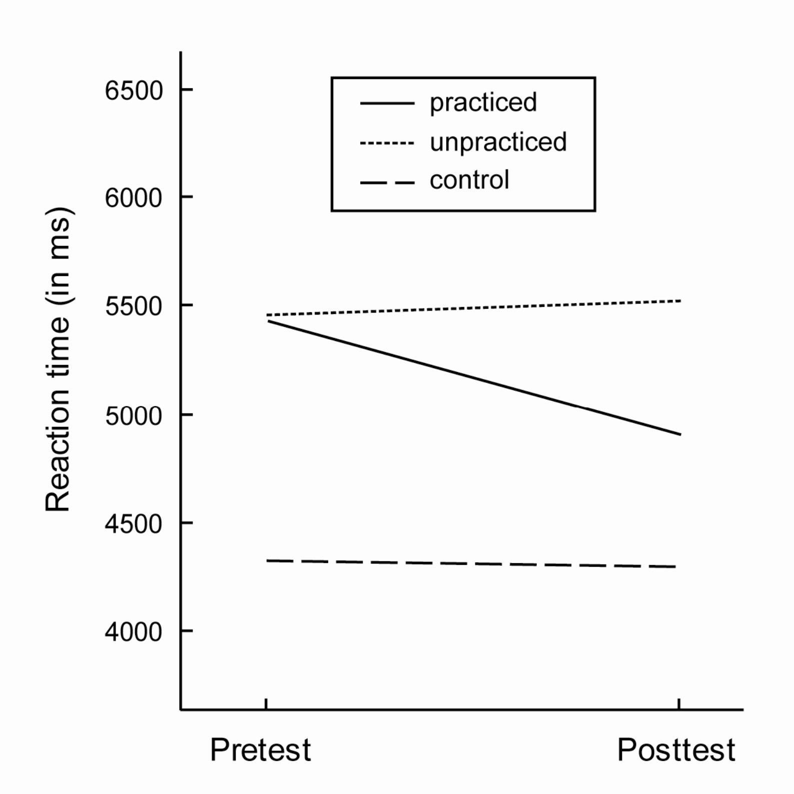

Figure 6.2. Addition problems in Grade 1. Mean reaction times (in ms) on the pretest and posttest for practiced, unpracticed, and control problems.

Test data. The mean reaction times for the three problem categories on the tests are presented in Figure 6.2. On the pretest, there was no significant difference between practiced and unpracticed problems. In contrast, control problems were answered much faster than the other two categories ($p < .001$); mean difference was more than a second. As can be seen in Figure 6.2, the practiced problems were the only category that improved from pretest to posttest. To evaluate if there was a significant improvement on any of the problem categories after practice, a repeated measures Anova was carried out with Time (pretest, posttest) and Problem (practiced, unpracticed, control) as within-subjects variables. The dependent variable was mean RT. Initially, Set (A or B) was included as between-subjects variable. The results showed no main effect for Set. There was one significant interaction effect with Set: Problem × Set ($p < .05$). This means that children who practiced set A scored differently on the three problem categories than children who practiced set B. However, because there was no significant interaction between Time, Problem, and Set, we decided to leave Set out of the final analysis. The repeated measures Anova with Time and Problem as within-subjects variables showed a main effect of Problem ($F(2, 74) = 26.90$, $p < .001$, $\eta_p^2 = .42$): control problems were faster than practiced ($p < .001$) and unpracticed problems ($p < .001$). Furthermore, there was a significant interaction effect between Time and Condition ($F(2, 74) = 4.01$, $p < .05$, $\eta_p^2 = .10$). Within-subject contrasts showed that the development from pretest to posttest on the practiced problems was significantly different from the development on the unpracticed ($p < .05$) and control problems ($p < .05$).

6.2.2.2 Intercorrelations

Table 6.4 shows descriptive statistics for all measures of academic achievement and cognitive processing and also for the two measures for improvement on the practice task. The mean RT for a problem on the pretest is included as an additional measure of mathematical ability. Note that some children had negative scores for improvement on the practice task, which means that there was actually no improvement. In fact, 11 % of the children had a negative score on the measure for improvement during practice and 37 % of the children had a negative score on the measure for improvement on the tests.

Table 6.4. Descriptive statistics for the measures in the addition experiment. Note that for the measures in the upper part of the table a higher score indicates a better performance, whereas for the measures in the lower part a higher score indicates a slower performance (all time measures are given in seconds).

	M	*SD*	Range
Math general	36.24	14.34	11 - 77
Addition	8.63	2.17	4 - 13
Subtraction	6.53	2.36	2 - 12
Reading speed	18.13	8.49	6 - 46
Digit span forward	5.16	1.29	2 - 8
Digit span backward	2.87	1.34	0 - 5
Learning names	7.08	3.18	2 - 16
Improvement practice	0.05	0.04	(-0.04) - (0.14)
Improvement tests	0.09	0.19	(-0.30) - (0.42)
Counting speed	13.01	3.86	7.2 - 22.7
RAN letters	0.80	0.21	0.44 - 1.36
RAN digits	0.84	0.19	0.55 - 1.46
RAN quantity	1.28	0.33	0.87 - 2.40
Mean RT pretest	5.07	1.64	2.79 - 8.67

Table 6.5 shows the intercorrelations among the measures of academic achievement. The first thing to notice is the absence of significant interactions between Mathematics general, Addition, and Subtraction. This was unexpected, because addition and subtraction are usually related and the general mathematics test treated both addition and subtraction. The significant correlation between Mathematics general and Reading speed confirms the finding that performance in reading and mathematics are related to each other (Jordan, Kaplan, & Hanich, 2002). However, from the last column of Table 6.5 we learn that the three measures of mathematical ability, especially Mathematics general and Addition, seem to predict performance on the pretest, whereas Reading speed does not.

Table 6.5. Intercorrelations among measures of academic achievement. Note that for the first four measures a higher score indicates a better performance, whereas a higher mean reaction time on the pretest indicates a slower performance.

	1.	2.	3.	4.	5.
1. Math general	–	.26	.24	.35*	-.48**
2. Addition		–	-.17	.19	-.43**
3. Subtraction			–	.23	-.33*
4. Reading speed				–	-.28
5. Mean RT pretest					–

* $p < .05$, ** $p < .01$

Table 6.6 shows the intercorrelations among the measures of cognitive processing. The high correlations between the different RAN scores were expected, just as the correlation between the score for Digit span forward and Digit span backward. Interesting is that children with higher scores for Learning names also have higher scores on the Digit span backward task and are faster on counting.

Table 6.6. Intercorrelations among measures of cognitive processing. Note that for measures 1 to 3 a higher score indicates a better performance, whereas for measures 4 to 7 a higher score indicates a slower performance.

	1.	2.	3.	4.	5.	6.	7.
1. Digit span forward	–	.36*	.16	-.15	.16	-.04	.18
2. Digit span backward		–	.45**	-.16	.02	-.04	-.10
3. Learning names			–	-.35*	-.16	-.27	-.19
4. Counting speed				–	.12	.07	.03
5. RAN letters					–	.81**	.51**
6. RAN digits						–	.67**
7. RAN quantity							–

* $p < .05$, ** $p < .01$

6.2.2.3 Correlations between cognitive processing and academic achievement

Table 6.7 shows the correlations between the measures of cognitive processing and the measures of academic achievement. Note that higher scores for counting speed, RAN, and mean RT on the pretest indicate slower performance. The table shows that Addition and Subtraction are not related to any of the cognitive measures. In contrast, children with higher scores for Mathematics general also have higher scores for Digit span backward and Learning names, and they are faster on counting. Reading speed is of course related to naming letters and digits. And finally, the mean RT on the pretest is related to Digit span backward and Counting speed.

Table 6.7. Correlations between measures of cognitive processing (rows) and measures of academic achievement (columns).

	Math general	Addition	Subtraction	Reading speed	Mean RT pretest
Digit span forward	.19	-.01	-.00	.14	-.18
Digit span backward	.42**	.06	.12	.20	-.37*
Learning names	.63**	-.05	.19	.28	-.23
Counting speed	-.45**	-.28	-.15	-.06	.45**
RAN letters	-.18	-.28	-.16	-.59**	.14
RAN digits	-.12	-.14	-.16	-.40*	.16
RAN quantity	.14	.03	-.18	-.04	.06

* $p < .05$, ** $p < .01$

6.2.2.4 Individual differences

The main question in this study was to find out whether individual characteristics of children predict their ability to learn arithmetic facts. For this reason, children performed a practice task with addition problems, and then we calculated how much they improved. In Table 6.8 the correlations between academic achievement and cognitive processing (rows) and the two measures for improvement on the practice task (columns) are presented. As explained earlier, we tried to control for the effect that slow children have a larger practice effect simply because they were slower to begin with. However, as can be seen in Table 6.8, this effect has by no means disappeared. Note that for the variables on the left side of the table a higher score indicates a better performance. To begin with the significant correlations, a higher score on Addition and on Digit span forward was related to a *lower* improvement during practice. And a higher score on Digit span backward was related to a *lower* improvement on the tests. The other correlations are not significant, but almost all of them are negative. The opposite is true for the variables on the right side of the table, where a higher score indicates a lower performance; here almost all of the correlations are positive, but none is significantly different from zero. To sum up, children who showed a large improvement on the practice task performed worse on the tasks for cognitive processing, which is of course contrary to expectations. The results suggest that the chosen improvement measures were inadequate to serve as an index for the ability to learn arithmetic facts. For this reason, we reanalyzed the data in two different ways. The first time we chose a logarithmic instead of a linear improvement measure. The second time we split the participants in one group with a large improvement and one group with a small or no improvement on the practice task and used ANOVA’s to analyze the data. However, all of these analyses basically led to the same results.

Table 6.8. Correlations between measures of academic achievement and cognitive processing (rows) and the two measures for improvement on the practice task (columns).

	Improvement practice	Improvement tests		Improvement practice	Improvement tests
Math general	-.13	-.29	Counting speed	.13	-.02
Addition	-.43**	-.02	RAN letters	.24	.24
Subtraction	-.11	-.18	RAN digits	.22	.32
Reading speed	-.29	-.20	RAN quantity	.15	.09
Digit span forward	-.49**	-.09			
Digit span backward	-.25	-.33*			
Learning names	.02	-.27			

* $p < .05$, ** $p < .01$

Another problem is that not all of the children seemed to benefit from practice, because as much as 37 % of the children did not show an improvement from pretest to posttest. For some part, this may be due to the small number of repetitions (12) of the problems during practice. However, this is probably not the whole story, because a much lower percentage of the children (11 %) showed no improvement during the practice sessions. The finding that some of the children improved during practice but not on the tests suggests that, because the posttest was placed at the end of the second individual test session, fatigue may have slowed down the participants' reaction times.

In Experiment 2, we made a number of modifications to the experimental design. (a) The dependent variable was accuracy within 4 seconds, instead of reaction time. (b) The number of repetitions was higher; 16 instead of 12. (c) The practice sessions were spread out over two weeks instead of two days. (d) The test sessions were on different days than the practice sessions. (e) There was not only a pretest and a posttest, but also a retention test (three weeks after the practice period). (f) There were more participants in order to calculate correlations more reliably.

6.3 Experiment 2: Multiplication

6.3.1 Method

6.3.1.1 Participants

The participants were 79 Grade 3 students (43 girls, 36 boys) from six Dutch schools for primary education. Mean age of the children was 8.7 years (*SD* = 0.53). Participants were selected with the same mathematics test as used in the addition experiment (De Vos, 1992): children performed 1 minute of additions, 1 minute of subtractions, and 1 minute of multiplications, with problems increasing in difficulty. Selected children had below-average to average scores on the multiplication problems, relative to the mean score in their school class. Participants' mean score was 10.5 multiplication problems (*SD* = 2.4) answered correctly in one minute. The most recent mathematics test of the pupil monitoring system (Cito, 2002) was administered about 4 months before the experiment was carried out. One school class did not participate in the pupil monitoring system. The normalized scores of the other 65 children showed that 7 participants scored in the lowest quartile, 15 in the second quartile, 26 in the third quartile, and 17 in the highest quartile of the normal population. In other words, although children with relatively low scores on multiplication were selected, the participants certainly did not have poor mathematical skills.

6.3.1.2 Academic achievement measures

Five measures of academic achievement were included as predictors for a learning effect in the practice task.

1. Mathematics general. This was the score from the mathematics test of the pupil monitoring system (Cito, 2002) and was obtained from the schools.

2-4. Addition / Subtraction / Multiplication. The score consisted of the number of addition / subtraction / multiplication problems answered correctly in one minute. The mathematics test (De Vos, 1992) was administered to a whole school class at the same time.

5. Orthographic knowledge. Orthographic knowledge was assessed with a paper and pencil test that consisted of 50 word pairs. One word of the pair was spelled correctly, the other word sounded the same (a homophone) but was spelled incorrectly. Children were asked to select the correct word by crossing out the incorrect version. The test was administered to a whole school class at the same time. The score consisted of the number of correct items.

6.3.1.3 Cognitive processing measures

Eleven measures of cognitive processing were included as predictors for a learning effect in the practice task. Procedures for **Digit span forward**, **Digit span backward**, and **Learning names** (measures 1-3) were the same as in the addition experiment.

4. Counting speed. Children were instructed to count forward from 34 to 47 and from 57 to 70. Responses were timed on a stopwatch from the first to the last item, measured to the nearest tenth of a second. The two counting scores were added together to obtain one score for counting speed. The intercorrelation between 34-47 and 57-70 was $r = .36$, $p < .01$.

5-7. Rapid automatized naming (RAN): letters / digits / quantity. The procedure for the RAN tasks was the same as in the addition experiment, with the exception that the task was not aborted after one minute. A stopwatch was used to measure the child’s naming time to the nearest tenth of a second. The score consisted of the time needed to name all 50 items.

8. Word span. The procedure and scoring method were the same as in the digit span forward task, but the sequences consisted of monosyllabic frequently used words.

9. Corsi block span. The Corsi block task (Corsi, 1972) is widely used to measure visual-spatial working memory. The apparatus consisted of nine identical blocks fastened to random positions on a board. To aid the experimenter, the numbers 1-9 were printed on the sides of the blocks facing the experimenter, invisible to the child. The experimenter pointed to a sequence of blocks at a rate of approximately one block per second. After the experimenter completed tapping the sequence, the child was asked to tap the blocks in the same order. The length of the block sequence was increased with 1 after two trials of the same length. The task started with a span length of two blocks and continued until the child made a mistake in both trials of the same span length. Scoring was identical to the digit span tasks; the score for the task consisted of the total number of sequences recalled correctly. Corsi block span has a test-retest reliability of .53 (Gathercole et al., 2004).

10. Mental rotation. Mental rotation was assessed using a subtest of a Dutch intelligence test for special education (Koornstra, Neuwahl, & Van Hoorn, 1979). The test was administered to a whole school class at the same time, together with the mathematics test and the orthographic knowledge test. The test had 10 problems; each problem consisted of a horizontal row with six symbols. The first symbol was the original symbol and was separated from the other five symbols with a vertical line. In the instruction, children were told that two of the five symbols were rotations of the original symbol and they had to cross out the three incorrect symbols. The incorrect symbols were rotations of the mirror image of the original symbol. The score consisted of the number of problems answered correctly.

11. Learning nonsense words. The procedure and scoring method were similar to the learning names task, but instead of learning associations between animals and names, children were asked to remember the meaning of “foreign” words. The foreign words were nonsense monosyllabic and bisyllabic words that sounded as valid Dutch words (e.g., “vels” or “molder”). Half of the monosyllabic foreign words were assigned to an arbitrary Dutch monosyllabic word and the other half were assigned to an arbitrary Dutch bisyllabic word; the same thing happened with the foreign bisyllabic words. In the first round the experimenter read the foreign words aloud and explained its meaning twice. In round 2 the words were read again and the child was asked whether it remembered the translation. If not, the experimenter explained the meaning of the nonsense word again. In round 3 the words were presented for the last time. The score for this task consisted of the number of correct answers in round 2 and 3.

6.3.1.4 Material and procedure of the practice task

Practice sessions. Two practice sets were constructed, consisting of 7 multiplication problems. The problems of set A were comparable in problem size to the problems of set B; see Table 6.9 for a list of the two sets. Half of the children practiced set A and the other half practiced set B. The problems were presented in a practice environment that was installed on the computers at school. The teacher made sure that students practiced two times a week in both practice weeks. A trial started with a problem printed in white on a black background. Children could enter their answer in an answer field under the problem. There was no time limit for answering. After typing in the answer, the children pressed ‘Enter’. If the answer was correct, a green check mark appeared in front of the answer field, in which the answer of the child was still visible. At the same time, the correct answer was printed next to the problem and the problem changed from white to green. If the answer of the child was incorrect, a red cross appeared in front of the answer field and the problem and correct answer were printed in red. The screen with problem and answer remained visible until the child pressed ‘Enter’ again and the next problem was presented. Reaction time and accuracy were registered by the computer. During the practice task, children practiced each problem 16 times. The material was built up from 16 blocks with 7 practice problems and 2 filler problems. Filler problems were relatively easy multiplication problems: tie problems and problems with 1, 5, or 10 as one of the factors. The order of the practice problems was pseudo-random: in consecutive trials there was never the same number on the same position in the problem (first addend, second addend, outcome). A practice session had 36 trials. The median time children needed to finish the first session was 10 minutes and the median time for the fourth session was 6 minutes. Although 104 children initially participated, 25 children were rejected from the analysis because they did not finish all four practice sessions (11), practiced five times instead of four (10), or because they typed in answers that were not numbers (4).

Table 6.9. Multiplication problems presented on the pretest, posttest, and retention test. During the practice sessions, half of the children practiced the problems of set A and the other half practiced the problems of set B.

Set A	Set B
2 × 6 = 12	3 × 4 = 12
9 × 2 = 18	8 × 2 = 16
8 × 3 = 24	4 × 6 = 24
4 × 7 = 28	9 × 3 = 27
8 × 4 = 32	4 × 9 = 36
6 × 8 = 48	6 × 7 = 42
6 × 9 = 54	7 × 8 = 56

Pretest, posttest, and retention test. Participants were tested on the problems of set A and set B. The procedure was similar to the tests in the addition experiment, but this time there was a time limit of 4 seconds. The time limit is expected to encourage the use of a retrieval strategy. Together with the presentation of the problem, four horizontal bars appeared at the bottom of the screen. Every second the rightmost bar ‘ticked away’ to indicate the time left to answer. Children were encouraged to answer within the time limit. As soon as the child answered, the experimenter pressed ‘Enter’ on a separate numeric keyboard and the problem disappeared. The experimenter then entered the answer of the child and pressed ‘Enter’ to start a new trial. If an answer was outside the time limit, it was labeled incorrect. All test problems were presented twice. The test started with 5 filler problems, after which 28 experimental trials were presented. Children needed about 5 minutes to finish a test.

6.3.1.5 General procedure

Each child was tested individually in three sessions of about 15 minutes, conducted on three different days. Testing took place in a quiet room at school. During the three individual test sessions, children performed all cognitive processing tasks. Furthermore, they performed the pretest, posttest, and retention test of the practice task. All children performed the tasks in the same order; see Table 6.10 for the general procedure. Between the pretest and the posttest, children practiced for two weeks (in four practice sessions) on a computer at school. Between the posttest and the retention test three weeks passed without practice.

Table 6.10. Tasks presented to the children in the three individual test sessions.

Tasks session 1	Tasks session 2	Tasks session 3
1) Pretest practice task	1) Posttest practice task	1) Retention test practice task
2) Digit span	2) RAN	2) Corsi block span
3) Counting speed	3) Learning nonsense words	3) Learning names
4) Word span		

6.3.1.6 Analysis of the practice task

Practice data. Filler trials were removed first. The accuracy scores during practice confirmed the assumption that the multiplication problems were relatively novel to the children, because many errors were made. In total, 27 % of the trials was incorrect. Accuracy for each of the 8 practice blocks was calculated (a practice block was defined as half of a practice session; in each practice block the practice problems were presented twice). The accuracy per block was used for the repeated measures analysis of the practice data, see Results. The median reaction time for correct responses was 7 seconds. Because participants made so many errors, we did not analyze reaction times for correct responses.

Test data. Filler trials were removed first. Then, accuracy scores were calculated for both categories (practiced, unpracticed) on each test (pretest, posttest, retention test). These scores were used for the repeated measures analysis of the test data, see Results. As a measure for improvement on the tests the absolute difference in accuracy on practiced problems between pretest and posttest was taken. There was no correlation between this measure for improvement and accuracy on the pretest.

6.3.2 Results

6.3.2.1 Practice task

Practice data. Figure 6.3 shows the mean number of items correct in the eight practice blocks. In each practice block all practice problems were presented twice. In contrast to the addition experiment there was no physical break between the practice blocks, so here a practice block was actually half of a practice session. To evaluate if accuracy improved during practice, a repeated measures Anova was carried out with Block (1 to 8) as within-subjects variable and accuracy as dependent variable. There was a main effect of Block ($F(7, 546) = 3.11$, $p < .01$, $\eta_p^2 = .04$), indicating that the improvement visible in Figure 6.3 was significant. Polynomial within-subjects contrasts showed a linear trend in the data of the practice blocks ($F(1, 78) = 11.39$, $p < .01$, $\eta_p^2 = .13$). There was no significant effect for a quadratic or a cubic trend. We could not find an explanation for the temporary decrease in accuracy in practice block 6, but overall it is clear that accuracy scores improved during practice. Improvement was moderate, however, because the mean number of correct items improved with only 1 problem from the first to the last practice block. The median time to finish one practice session decreased from 10 minutes for the first session to 6 minutes for the last session. This suggests that children also improved in reaction time during practice.

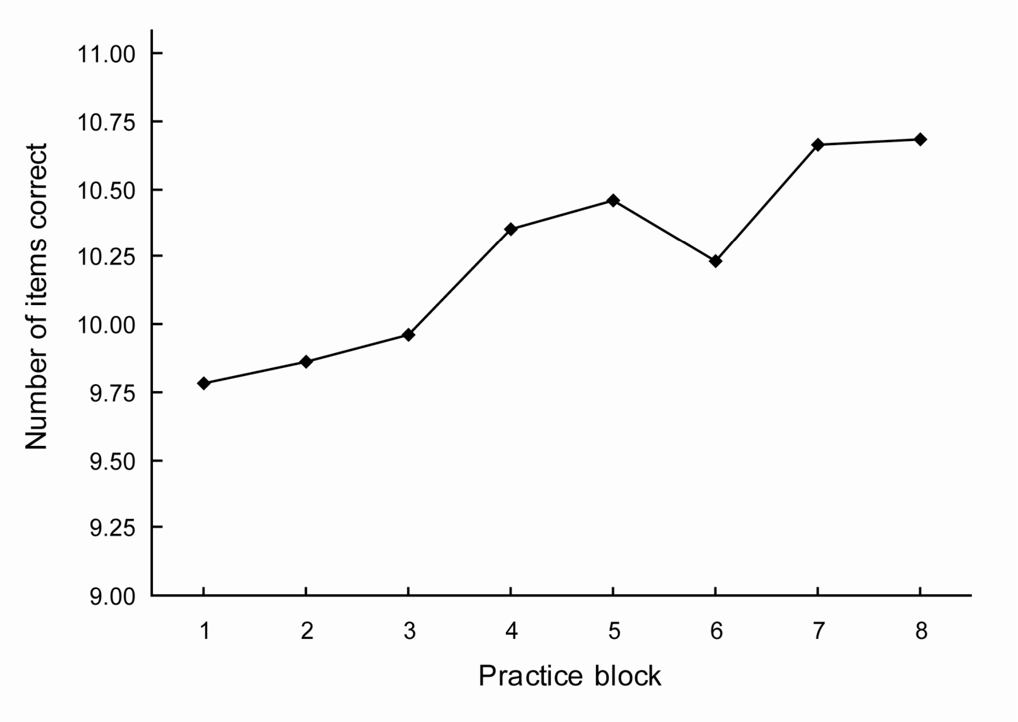

Figure 6.3. Multiplication problems in Grade 3. Mean number of items correct in the eight practice blocks. Maximum score is 14.

Test data. Figure 6.4 shows the scores for accuracy on practiced and unpracticed problems on the three test occasions; the data is presented separately for children who practiced set A and children who practiced set B. To evaluate if there was an improvement on any of the problem categories after practice, a repeated measures Anova was carried out with Time (pretest, posttest, retention test) and Problem (practiced, unpracticed) as within-subjects variables. Set (A or B) was included as between-subjects variable. The results showed a main effect of Time ($F(2, 154) = 124.53$, $p < .001$, $\eta_p^2 = .62$): mean scores improved from pretest (3.3 items correct) to posttest (6.0 items correct) and then decreased a little in the retention period (5.8 items correct on the retention test). There was also a main effect of Problem ($F(1, 77) = 37.22$, $p < .001$, $\eta_p^2 = .33$): scores were higher for practiced problems (5.6 items correct) than for unpracticed problems (4.5 items correct). Furthermore, there was a significant interaction effect between Time and Problem ($F(2, 154) = 25.23$, $p < .001$, $\eta_p^2 = .25$). On the pretest, there was no significant difference between practiced and unpracticed problems, but within-subjects contrasts showed that practiced problems improved more from pretest to posttest than unpracticed problems ($p < .001$). The same was true from pretest to retention test: practiced problems improved more than unpracticed problems ($p < .001$). Finally, a significant interaction between Time, Problem, and Set was found ($F(2, 154) = 5.56$, $p < .01$, $\eta_p^2 = .07$). However, as can be seen in Figure 6.4, the development in scores for children who practiced set A was very similar to the development in scores for children who practiced set B. For both sets, practiced problems improved more than unpracticed problems from pretest to posttest ($p < .001$). For children who practiced set B the difference in improvement between practiced and unpracticed problems was also significant from pretest to retention test ($p < .001$). In contrast, for children who practiced set A the difference in improvement between practiced and unpracticed problems was no longer significant from pretest to retention test ($p = .43$).

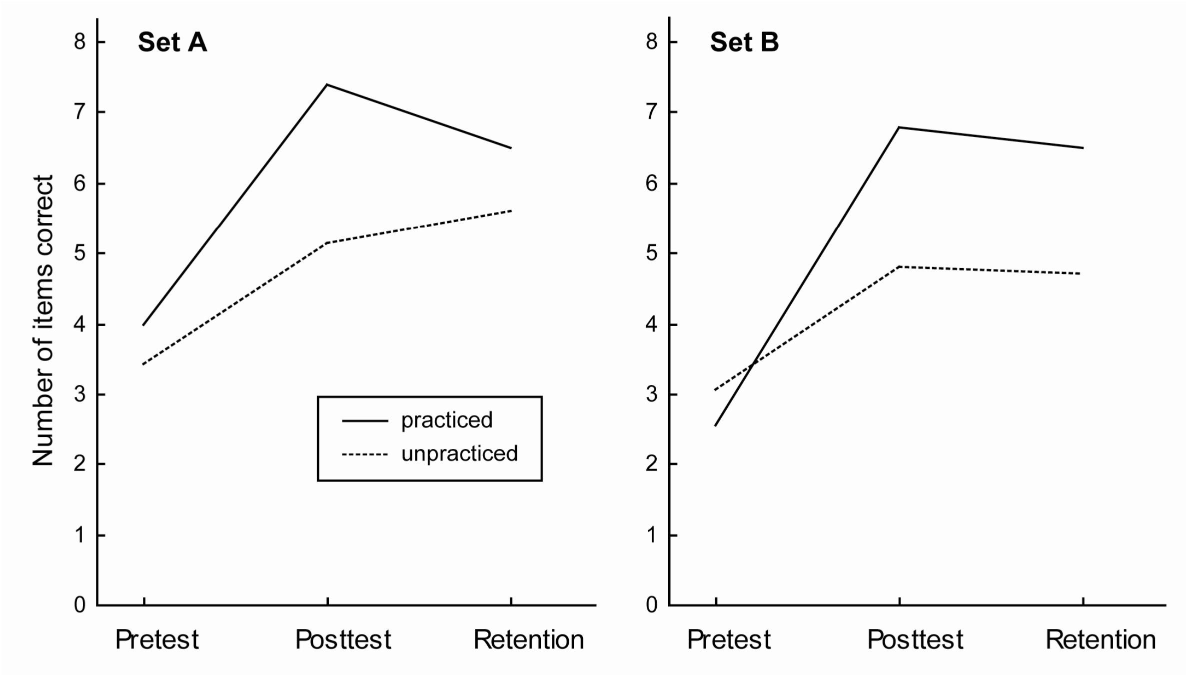

Figure 6.4. Multiplication problems in Grade 3. Mean number of items correct on pretest, posttest, and retention test, presented separately for children who practiced set A and children who practiced set B. Maximum score is 14.

6.3.2.2 Intercorrelations

Table 6.11 shows descriptive statistics for all measures of academic achievement and cognitive processing and also for the measure for improvement on the practice task. Again, the range of the improvement measure shows that not all children improved, but this time the percentage of children with no improvement on the tests was much lower (10 %). Accuracy on the pretest (maximum score is 28) is included as an additional measure of mathematical ability. Note that although the Learning nonsense words task was designed to be similar to the Learning names task, children found Learning words much more difficult. Also note that children were much slower on RAN quantity than on the other two RAN tasks. During the RAN quantity task we observed that, especially at the beginning of the task, children needed to count the groups with 4 and 5 triangles. This observation suggests that performance on the RAN quantity task relies in part on the ability to subitize small numbers.

Table 6.11. Descriptive statistics for the measures in the multiplication experiment. Note that for the last four measures a higher score indicates a slower performance (all time measures are given in seconds).

	M	*SD*	Range
Math general	72.82	8.69	50 - 95
Addition	16.42	3.65	8 - 27
Subtraction	13.16	3.98	7 - 25
Multiplication	10.51	2.39	6 - 15
Orthography	43.80	5.37	16 - 50
Digit span forward	6.38	1.09	4 - 10
Digit span backward	3.62	1.05	2 - 6
Word span	5.61	1.21	3 - 10
Block span	7.92	1.68	5 - 12
Mental rotation	6.27	3.01	0 - 10
Learning names	14.23	3.81	4 - 23
Learning words	7.27	3.39	1 - 15
Improvement tests	3.77	2.53	(-1) - (10)
Accuracy pretest	6.62	3.74	0 - 14
Counting speed	22.24	5.21	12.7 - 45.1
RAN letters	31.52	6.27	20.9 - 54.5
RAN digits	31.85	5.53	18.9 - 54.8
RAN quantity	52.21	9.93	33.4 - 79.2

Table 6.12 shows the intercorrelations among the measures of academic achievement. This time, all measures for mathematical ability were correlated to each other. The score for orthographic knowledge, which is a control variable, was not related to any of the mathematical scores.

Table 6.12. Intercorrelations among measures of academic achievement.

	1.	2.	3.	4.	5.	6.
1. Math general	–	.25*	.38**	.35**	.15	.27*
2. Addition		–	.65**	.39**	.07	.30**
3. Subtraction			–	.49**	.05	.47**
4. Multiplication				–	.19	.52**
5. Orthography					–	.08
6. Accuracy pretest						–

* $p < .05$, ** $p < .01$

Table 6.13 shows the intercorrelations among the measures of cognitive processing. There is no significant correlation between Digit span forward and Digit span backward. However, there is a correlation between Digit span forward and Word span, which are two tasks that differ only with respect to the type of stimuli. The two visual memory tasks, Block span and Mental rotation, are related. Learning Names and Learning words are not. And again, there are correlations between the different RAN tasks. The remaining significant correlations are between Digit span backward and Block span, Mental Rotation, Counting speed, and RAN quantity, and also between Block span and RAN quantity, but these correlations are rather low.

Table 6.13. Intercorrelations among measures of cognitive processing. Note that for measures 8 to 11 a higher score indicates a slower performance.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Digit span forward	–	.12	.51**	.04	.11	.25*	-.02	-.18	-.20	-.08	-.10
2. Digit span backward		–	.05	.24*	.26*	.20	.21	-.23*	-.15	-.11	-.28*
3. Word span			–	.09	.03	.20	-.10	-.04	.01	.06	.10
4. Block span				–	.34**	-.02	-.13	-.18	.00	-.01	-.22*
5. Mental rotation					–	.16	.02	-.07	.04	.15	.00
6. Learning names						–	.12	-.09	-.16	-.09	-.07
7. Learning words							–	.08	-.20	-.04	.02
8. Counting speed								–	.39**	.34**	.36**
9. RAN letters									–	.56**	.49**
10. RAN digits										–	.35**
11. RAN quantity											–

* $p < .05$, ** $p < .01$

6.3.2.3 Correlations between cognitive processing and academic achievement

Table 6.14 shows the correlations between the measures of cognitive processing and the measures of academic achievement. Note that higher scores for counting speed and RAN indicate slower performance. Again, the correlations are not very high, but the table shows that the cognitive processing tasks with the most significant correlations with the measures for mathematical ability are Digit span forward, Digit span backward, Counting speed, and RAN. The most striking result is that Digit span forward is significantly correlated with four of the measures of mathematical ability, whereas Word span is correlated with none of these measures. This suggests that the relationship between verbal short-term memory span and mathematical ability is domain-specific.

Table 6.14. Correlations between measures of cognitive processing (rows) and measures of academic achievement (columns).

	Math general	Addition	Subtraction	Multiplication	Orthography	Accuracy pretest
Digit span forward	.35**	.31**	.29*	.26*	.13	.21
Digit span backward	.35**	.24*	.21	.17	.04	.27*
Word span	.16	.21	.19	.12	.05	.17
Block span	.20	.16	.02	-.05	.06	.03
Mental rotation	.31*	-.08	-.07	.04	-.05	-.03
Learning names	.12	.05	.20	.07	.20	.04
Learning words	-.13	-.23*	-.04	-.06	-.08	.00
Counting speed	.02	-.24*	-.30**	-.17	.10	-.24*
RAN letters	-.03	-.28*	-.23*	-.07	-.04	-.03
RAN digits	.32*	-.33**	-.05	-.06	.06	.02
RAN quantity	-.09	-.31**	-.28*	-.10	-.25*	-.10

* *p* < .05, ** *p* < .01

6.3.2.4 Individual differences

Table 6.15 shows the correlations between academic achievement and cognitive processing (rows) and the measure for improvement on the practice task (columns). There is only one significant correlation, between Learning words and the measure for improvement, but this correlation is very low. Based on the results in Table 6.15, we conclude that it is not possible to predict children's improvement on the multiplication practice task from their scores on the tasks for cognitive processing.

Table 6.15. Correlations between measures of academic achievement and cognitive processing (rows) and the measure for improvement on the tests (columns).

	Improvement tests		Improvement tests		Improvement tests
Math general	-.02	Digit span forward	-.00	Counting speed	-.01
Addition	-.07	Digit span backward	-.08	RAN letters	-.00
Subtraction	.00	Word span	-.06	RAN digits	.04
Multiplication	.00	Block span	.01	RAN quantity	.08
Orthograpy	.01	Mental rotation	-.00		
Accuracy pretest	-.12	Learning names	.14		
		Learning words	.23*		

* *p* < .05, ** *p* < .01

6.4 Discussion

In two experiments, primary school children practiced simple arithmetic problems. In the first experiment, Grade 1 children practiced a small set of addition problems. Before and after practice, they were tested on three problem categories: 1) practiced problems, 2) unpracticed problems, comparable in difficulty level to the practiced problems, 3) control problems, which were unpracticed addition problems that could easily be solved by counting. The results of the tests showed a significant improvement in speed for the practiced problems, but not for the other two problem categories, which indicates that participants had learned from practice. Apart from the practice task, children performed four academic achievement tasks and seven cognitive processing tasks. The first question we addressed was whether scores on cognitive processing tasks were related to measures of mathematical achievement. The results showed a significant correlation between performance on Digit span backward and Counting speed on the one hand and performance on two measures of mathematical achievement (a general mathematics test and the mean reaction time on the pretest) on the other hand. Also, children with higher scores for learning associations between a picture of an animal and a relatively arbitrary name performed better on the general mathematics test. Scores for Digit span forward and Rapid automatized naming were not related to any of the measures for mathematical achievement. The second question we addressed was whether performance on the cognitive processing tasks could predict individual learning effects on the practice task. Therefore, two measures for improvement on the practice task were calculated and correlational analyses with the cognitive processing scores were carried out. The results showed two significant correlations, but these were against expectations: children with lower scores for Digit span forward improved most during practice and children with lower scores for Digit span backward improved most on the tests. These results suggest that we did not succeed in controlling for the effect that slow children had a larger practice effect simply because they were slower to begin with.

In the second experiment, Grade 3 children practiced for two weeks on a small set of multiplication problems. Before and after practice, they were tested on practiced and unpracticed problems. The results of the tests showed an improvement on accuracy from pretest to posttest that was significantly higher for practiced than for unpracticed problems, which indicates that participants had learned from practice. Apart from the practice task, children performed five academic achievement tasks and eleven cognitive processing tasks. Correlations between cognitive processing scores and mathematical achievement were not very high, but can be summarized as follows. Starting with the two verbal short-term memory tasks: a significant correlation with almost all measures of mathematical achievement was found for Digit span forward, but not for Word span. The verbal working memory task, Digit span backward, was also related to several mathematical achievement scores. For the visual-spatial tasks the only significant correlation was between Mental rotation and the general mathematics score. For the association tasks the only significant correlation was between Learning words and addition. Counting speed was again related to several mathematical achievement scores. And finally, there were also some relations between the scores for rapid naming and the mathematical achievement scores (mostly addition and subtraction). To answer the question whether performance on the cognitive processing tasks could predict individual learning effects on the practice task, a measure for improvement on the practice task was calculated. Correlational analyses between the measure for improvement and the cognitive processing scores were carried out. The results showed only one significant correlation, which was between Learning words and the improvement measure, but this correlation was very low. Therefore, we conclude that it is not possible to predict children's improvement on the multiplication practice task from their scores on the tasks for cognitive processing.

The results from the two experiments showed that performance on short-term memory, working memory, counting speed, and rapid naming tasks did not predict the children's ability to learn arithmetic facts in a learning task. Even though the literature on children with mathematical difficulties clearly shows that links exist between cognitive processing and mathematical ability, the present study found no evidence for associations between cognitive processing and arithmetic fact learning. How do we explain these results? First of all, it is possible that the dependent measures in the study were not appropriate. Apparently, it is not easy to choose an adequate dependent measure for this type of research, in which individual differences between children are related to their improvement on a mathematical achievement task. The dependent measure in the addition experiment was reaction time, but this measure suffered from a scaling effect that was not easy to overcome. In contrast, the dependent measure in the multiplication experiment – accuracy with a limited time for responding – did not seem vulnerable to a scaling effect. For this reason, a measure such as used in the multiplication experiment, which combines speed and accuracy, seems most promising for future research. On the other hand, such a measure is of course not very sensitive to small changes in improvement, which may explain why almost no correlations with the cognitive processing scores were found in the multiplication experiment. Furthermore, it is difficult to choose an adequate time limit, because this depends on the level of mathematical development of the children.

It is of course also possible that the ability to learn arithmetic facts is not related at all to measures of cognitive processing. In other words, this would mean that arithmetic fact learning is a unique ability that cannot readily be predicted from working memory or short-term memory capacity, counting speed or rapid naming. This view is not in accordance to other research that has shown that various links exist between cognitive processing and mathematical ability. However, other research has mainly focused on children with severe mathematical problems and perhaps it is not possible to extrapolate findings in children with specific mathematical disabilities to children with average mathematical abilities. Further research is needed to study the relationship between mathematics and cognitive processing in normal children.

Another issue is whether short time practice experiments provide a good model for the development of arithmetic fact learning. The results showed that most children learned from practice, so it is probably not necessary to have children practicing problems many more times to study improvement on arithmetic facts. Furthermore, it seems valuable to study children's ability to learn in a short period of time, because we would not be surprised if this has a predictive value for their further mathematical development. There are actually very few studies who used a learning task design to study individual differences in mathematics.

In the present study we introduced two tasks that measured the ability to form associations: Learning names and Learning words. Based on the results, we are not sure whether arithmetic fact learning is related to a more general ability to form associations. Indeed, there was a significant correlation between learning "foreign" words and the improvement on the multiplication practice task, but this correlation was low. Furthermore, the scores for Learning words were not related to the scores for Learning names and Learning words was also much more difficult than Learning names. This suggests that the correlation between Learning words and improvement on multiplication problems may have more to do with general intelligence than with learning associations. Nevertheless, we think that it is interesting to include a task like Learning names in future research. In the addition experiment Learning names was found to be related to both Digit span backward and Counting speed and in the multiplication experiment it was found to be related to Digit span forward. Further research could investigate the nature of these relationships.

As explained earlier, apart from the practice tasks we also included some general measures of arithmetical achievement. The correlational analyses between scores for cognitive processing and scores for mathematical achievement yielded some interesting results. In the addition experiment, Digit span backward and Counting speed were related both to the general mathematics achievement score and to performance on the pretest, which was actually a measure for addition. Furthermore, there was a significant correlation between Learning names and general mathematical achievement.

In the multiplication experiment, the first thing to notice is that the two measures for multiplication (multiplication on a one-minute test and performance on the pretest) had less significant correlations with the measures for cognitive processing than the other three measures for mathematical achievement (mathematics general, addition, subtraction). This difference may indicate that the nature of multiplication fact learning is different from addition fact learning. However, another explanation is that multiplication was relatively new to the children in this study, and therefore individual differences were not yet reliably developed. Indeed, the standard deviation for the scores on the one-minute multiplication test was lower than the standard deviation for the scores on the one-minute addition and subtraction tests. This restriction in range may have caused less significant correlations with the scores for cognitive processing.

Turning to the other measures for mathematical achievement – mathematics general, addition, and subtraction – it seems that Digit span forward, Digit span backward, Counting speed, and Rapid naming are all in some way related to mathematical achievement. Of course the correlations were low and we cannot say anything about causation. However, it is interesting that not only correlations between mathematics and regularly used tasks as digit span were found, but also between mathematics and rapid automatized naming. Research on the relation between rapid naming and mathematics is scarce, and it seems promising to explore this relationship further. Finally, an interesting result is that Digit span forward was significantly correlated to general mathematics, addition, subtraction, and also multiplication, whereas Word span was correlated with none of these measures. Digit span forward and Word span were similar tasks that differed only with respect to the type of stimuli. Furthermore, the mean scores of the children did not indicate that Word span was much more difficult than Digit span. Therefore, the results suggest that the relationship between verbal span and mathematical ability is domain-specific.

In conclusion, we were not able to explain individual learning effects on the practice task from differences between children on measures of cognitive processing. What we did found was that Digit span forward, Digit span backward, Counting speed and Rapid naming seem to be related to mathematical ability. Furthermore, the relationship between verbal short-term memory span and mathematical ability seems to be domain-specific.

6.5 Acknowledgments

We are grateful to all the students and teachers who participated in this study and to Baukje Hijlkema and Janet Hospers for their help in collecting the data.

7

General discussion

The background for the research described in this thesis is that children in primary school do not always succeed in gaining efficiency in solving simple arithmetic problems. Lacking efficiency in simple arithmetic is generally regarded as a serious problem, because fluent solving of simple arithmetic problems is a prerequisite for solving more complex problems. With this research, we hope to contribute to finding an explanation why some children seem to learn arithmetic facts almost automatically and other children keep struggling with basic addition and multiplication. The current studies have revealed much information about automatization and learning simple arithmetic. However, explaining individual differences in the ability to learn arithmetic facts was less successful. In this general discussion the findings of the previous chapters are summarized and discussed.

7.1 Practice effects and transfer effects

One of the most obvious conclusions that can be drawn from Chapter 4, 5, and 6 is that practice in simple arithmetic leads to better performance. Of course, this is not a new idea, since it is one of the basic principles of education. Nevertheless, it is important to realize that frequent repetition of simple arithmetic problems enhances the performance of children in a very effective way. Also important is that in general no differences were found between the learning effects of children with poor mathematical skills and children with good mathematical skills. Therefore, it appears that all children benefited from practice. Moreover, it does not necessarily have to take much time to practice simple arithmetic problems: short practice sessions on a regular basis might improve the students' performance substantially.

In education, one hopes that children are able to generalize knowledge from one domain to another domain. For instance, it is not sufficient that children know how to do a division, because they also have to recognize which situations in real life ask for a division procedure. Simple arithmetic facts have the interesting property that they are related to each other in many ways. Moreover, in contrast to for instance topographical facts, it is not always necessary to know arithmetic facts by heart, because there are many backup procedures. For instance, '9 × 8' can be transformed to the easier problem '[10 × 8] - 8'. The foregoing example illustrates that arithmetic facts can be calculated from each other. Therefore, it is possible that practicing certain arithmetic problems will lead to a transfer of knowledge to other arithmetic problems.

In Chapter 4 and Chapter 5, transfer effects in simple arithmetic were studied. In Chapter 4, a transfer effect to regular addition problems was found for children who practiced with the method of choosing between two possible answers. In Chapter 5, it was found that the commutative property of additions and multiplications helps children to learn simple addition and multiplication. When children practiced for example ‘6 + 3’, they also improved on ‘3 + 6’. And when they practiced ‘7 × 4’, they also improved on ‘4 × 7’. These findings suggest that children represent commutative problems as a single fact in long-term memory, and not as two separate facts. Children did not improve on related problems in which one of the operands of the original problem was increased with one unit, even when only a small step was needed to solve the problem (e.g., 6 + 4 = [6 + 3] +1).

A good model to explain the findings in Chapter 4 and 5 is the Identical Elements (IE) model of arithmetic fact representation (Rickard, 2005; Rickard & Bourne, 1996; Rickard, Healy, & Bourne, 1994). According to this model, there is a single long-term memory node for problems consisting of the same numerical elements (i.e., operands and answer), regardless of operand order. The IE model was developed for existing arithmetic fact knowledge in adults, but the present results show that it can also explain learning effect in children. Looking at students’ performance in simple arithmetic from the viewpoint of the IE model can be informative about which generalizations children find easy and which are difficult. For instance, the inverse relationships between addition and subtraction and between multiplication and division are probably not intuitive for children. Consequently, a large investment needs to be made in explaining conceptual knowledge about inverse relationships in mathematics education.

7.2 Basic numerical processing

Ever since the discovery of the SNARC effect (Dehaene, Bossini, & Giraux, 1993; Dehaene, Dupoux, & Mehler, 1990), there has been a huge interest in basic numerical processing. Research over the past decades has made clear that several robust effects occur in number processing. The SNARC effect refers to the finding that small numbers facilitate left responses, whereas larger numbers facilitate right responses. The distance effect refers to the finding that numerically close numbers are more difficult to compare than numbers further apart. Finally, the size effect refers to the finding that at the same numerical distance, performance in number comparison decreases as the size of the numbers increases. Together, these three effects gave rise to the idea of an internal representation of numbers as if placed on a compressed number line, with small numbers on the left side and large numbers on the right side. This mental number line is logarithmic in nature, because the perceived distance between small numbers is larger than the perceived distance between larger numbers. As is discussed in detail in Chapter 2, the assertion that the SNARC effect is caused by the representation of numbers on a left-to-right mental number line has been heavily debated, just as the idea that the left-to-right direction of the SNARC effect is directly related to the direction of writing in Western cultures. Nevertheless, the discovery of the SNARC effect has inspired many researchers to study basic numerical processing and much progress has been made over the past years.

Given the large number of studies on the SNARC effect it is remarkable how few studies have included children as participants. It is of course convenient to study the SNARC effect in psychology students, but if we want to understand the origin of this effect it is essential to study the SNARC effect also in children. For this reason, I believe that the study in Chapter 2 makes an important contribution to the theoretical discussion on the SNARC effect and the development of automaticity in number processing. Studying the SNARC effect may even be important for understanding automatic processes outside the domain of mathematical cognition, for instance in the case of automaticity in the allocation of attention.

Recently, an interest in basic numerical processing is also starting to come from research on mathematical difficulties. There have been a few studies that suggest that children and adults with mathematical difficulties are impaired on accessing number magnitude from Arabic numerals (Bachot, Gevers, Fias, & Roeyers, 2005; Rouselle & Noël, 2007; Rubinsten & Henik, 2005). Research in this direction may be fruitful in understanding the origins of mathematical difficulties. Another line of research suggests that the link between numerical estimation and mathematical skills (see Chapter 3) is mediated by the ability to move away from the natural, logarithmic way of representing number magnitudes, and represent numbers in a linear way. Children with a linear representation of numbers were found to have better mathematical skills than children with a logarithmic representation (Booth & Siegler, 2006). An early screening of basic number processing skills and the ability to represent number magnitudes in a linear way might prove a powerful means to predict mathematical problems later in education.

7.3 Working memory and rapid automatized naming

A subject that has become increasingly important in the research on children's learning of mathematics is working memory. Over the last decades, it has been shown that poor working memory limits a child's performance in mathematics in a substantial way. In Chapter 6, the relationships between cognitive processing and mathematical achievement were studied. Although it was not possible to predict individual learning effects on two arithmetic practice tasks from scores for cognitive processing, several correlations were found between cognitive processing and mathematical achievement. Interestingly, not only correlations between mathematics and regularly used tasks as digit span forward (short-term memory) and digit span backward (simple working memory) were found, but also between mathematics and rapid automatized naming, which is a measure of the ease of retrieving phonological representations from long-term memory. Research on the relation between rapid automatized naming and mathematics is scarce, and it seems promising to explore this relation further. If children with mathematical problems are slow on digit naming, this could indicate that they are impaired on basic numerical processing. Furthermore, slow digit naming could be a symptom of a more general problem in the storage and access of numerical material in long-term memory, including basic arithmetic facts.

7.4 Automatization in simple arithmetic

In the present thesis, a broad approach was taken towards automatization in simple arithmetic. Automatization was defined as gaining efficiency in solving simple arithmetic problems, which means that problems are solved faster and less errors are made. As is argued in Chapter 5, the selection of an appropriate dependent variable in the experiments is related to the level of skill in the participants. If the majority of the answers is correct on the pretest, reaction times are suitable to measure performance. If this is not the case, reaction times are not adequate and accuracy needs to be chosen as performance measure. Both in Chapter 5 and Chapter 6, reaction times were selected as the dependent variable for the addition experiment in Grade 1, whereas accuracy (within a certain time limit) was selected as the dependent variable for the multiplication experiment in Grade 3. Even though different measures were used, the results in Chapter 5 were very similar for children in Grade 1 and Grade 3. This shows that different approaches can lead to the same result.

The findings in Chapter 6 are more complex. In this study, individual differences on cognitive processing measures were related to children's improvement on a mathematical achievement task. In this design it was crucial that individual mathematical improvement scores were very precise. Which measure for performance is suitable in this situation, reaction time or accuracy? Reaction time measures are sensitive to small improvements, but they are also extremely variable both within and between children of this young age. To overcome this large variance a good method of rejecting outliers needs to be chosen, which of course requires a sufficient number of reaction times. Also, it was observed that if individual improvement measures are calculated based on reaction times, the children with the lowest solution times will generally improve most, which is a scaling problem that is very difficult to overcome. The alternative to using reaction times is accuracy, but this measure is not without complications either. For practical reasons it is necessary to choose a time limit for answering, because otherwise children who do not have efficient retrieval or procedural strategies available might spend a long time to answer. Consequently, the experiment would take very long and this could demotivate the participants.

Although a measure like accuracy with a limited time for responding seems to capture the essence of automaticity in a good way, because it combines reaction time and accuracy, the duration of the time period for responding is of course more or less arbitrary. Which time limit is adequate depends on the level of mathematical development of the children: a time limit that is too short might cause a floor effect, whereas a time limit that is too long might cause a ceiling effect. In both cases the sensitivity of the accuracy measure will decrease.

The results in Chapter 6 do not only show that it can be difficult to select a suitable measure for performance, but also that it is not easy to model the acquisition of arithmetic facts in a short learning task experiment. Several correlations were found between individual scores on measures of cognitive processing and mathematical ability at the time of the pretest, but these cognitive processing measures were not related to the improvement on the practice task. In other words, the relations were visible only when looking at the level of skill in mathematics that had developed over a number of years and not when looking at the ability to learn arithmetic facts in a short period of time. Nevertheless, I think that it would be interesting to use learning task designs more often in educational research. Measuring what a child is able to learn in a short period of time may be indicative of the child's success in education.

7.5 Conclusions

The main conclusions from the studies in this thesis can be summarized as follows. Seven-year-olds represent number magnitudes in a way similar to that of adults, because they show a SNARC effect in a magnitude judgment task (Chapter 2). When perceiving Arabic numerals, children have developed automatic access to magnitude information by around 9 years of age (Chapter 2). Numerical estimation is related to mathematical ability, especially in younger children (Chapter 3). Choosing between two alternative answers is an effective practice method for learning addition facts (Chapter 4). Children probably have one representation in long-term memory for commutative addition and multiplication facts (Chapter 5). The Identical Elements model is a good model to explain children's transfer effects in addition and multiplication (Chapter 4 and 5). Working memory, short-term memory, counting speed, and rapid naming are related to mathematical ability (Chapter 6). And finally, practice in simple arithmetic leads to better performance (Chapter 4, 5, and 6).

References

Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, *67*, 21-38.

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology, 100*, 48-66.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport*, *16*, 1769-1773.

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In C. J. Brainerd, R. Kail, & J. Bisanz (Eds.), *Formal methods in developmental research* (pp. 302-338). New York: Springer-Verlag.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75-106.

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in Grades 1-6. *Journal for Research in Mathematics Education*, *26*, 396-421.

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology Science*, *47*, 172-183.

Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia*, *36*, 731-735.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*, 5-28.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, *Vol. 8* (pp. 47-89). New York: Academic Press.

Baroody, A. J. (1999a). Children's relational knowledge of addition and subtraction. *Cognition and Instruction*, *17*, 137-175.

Baroody, A. J. (1999b). The roles of estimation and the commutativity principle in the development of third graders' mental multiplication. *Journal of Experimental Child Psychology*, *74*, 157-193.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp.1-33). Mahwah, NJ: Erlbaum.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, *74*, 286-308.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test*. [Revision Amsterdam Child Intelligence Test]. Lisse: Swets & Zeitlinger.

Blöte, A. W., Klein, A. S., & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction*, *10*, 221-247.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*, 189-201.

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, *20*, 607-618.

Brus, B. T., & Voeten, M. J. M. (1973). *Eén-Minuut-Test*. [One minute reading test]. Nijmegen: Berkhout.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, *65*, 1-24.

Butterworth, B., Marchesini, N., & Girelli, L. (2003). Basic multiplication combinations: Passive storage or dynamic reorganization? In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp.189-202). Mahwah, NJ: Erlbaum.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 109-123.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*, 1-44.

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121-164.

Campbell, J. I. D., & Austin, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory & Cognition*, *30*, 988-994.

Campbell, J. I. D., & Clark, J. M. (1989). Time course of error priming in number-fact retrieval: Evidence for excitatory and inhibitory mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 920-929.

Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition*, *34*, 633-647.

Carpenter, T. P. (1986). Conceptual knowledge as a foundation for procedural knowledge: Implications from research on the initial learning of arithmetic. In J. Hiebert (Ed.), *Conceptual procedural knowledge: The case of mathematics* (pp. 113- 132). Hillsdale, NJ: Erlbaum.

Cito (1992). *Drie-Minuten-Toets* [Three-Minute Test]. Arnhem: Cito.

Cito (2002). *Leerlingvolgsysteem Rekenen-Wiskunde*. [Student monitoring system: Mathematics]. Arnhem: Cito.

Cito (2005). *Leerlingvolgsysteem Rekenen-Wiskunde*. [Student monitoring system: Mathematics]. Arnhem: Cito.

Cooney, J. B., Swanson, H. L., & Ladd, S. F. (1988). Acquisition of mental multiplication skill: Evidence for the transition between counting and retrieval strategies. *Cognition and Instruction*, *5*, 323-345.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Unpublished doctoral dissertation, McGill University, Montreal, Canada.

Crites, T. (1992). Skilled and less skilled estimators' strategies for estimating discrete quantities. *The Elementary School Journal*, *92*, 601-619.

Cumming, J. J., & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition*, *5*, 149-180.

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, *15*, 189-202.

Dagenbach, D., & McCloskey, M. (1992). The organization of arithmetic facts in memory: Evidence from a brain-damaged patient. *Brain and Cognition*, *20*, 345-366.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371-396.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219-250.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626-641.

Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters, and numbers by normal children. *Cortex*, *10*, 186-202.

Deschuyteneer, M., & Vandierendonck, A. (2005). Are "input monitoring" and "response selection" involved in solving simple mental arithmetical sums? *European Journal of Cognitive Psychology*, *17*, 347-370.

De Vos, T. (1992). *Tempo Test Rekenen*. [Speeded arithmetic test]. Nijmegen: Berkhout.

Dowker, A. (2003). Young children's estimates for addition: The zone of partial knowledge and understanding. In A.J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 243-265). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments - An additive-factors analysis. *Memory & Cognition*, *8*, 612-622.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year-old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*, 1-17.

Fias, W., Brysbaert, M., Geypens F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, *2*, 95-110.

Fias, W., & Fischer, M. H. (2005). Spatial representation of numbers. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 43-54). New York: Psychology Press.

Fias, W., Lauwereyns, J., & Lammertyn, J. (2001). Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Cognitive Brain Research*, *12*, 415-423.

Fischer, M. H. (2003). Spatial representations in number processing - Evidence from a pointing task. *Visual Cognition*, *10*, 493-508.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience*, *6*, 555-556.

Fischer, M. H., Warlop, N., Hill, R. L., & Fias, V. (2004). Oculomotor bias induced by number perception. *Experimental Psychology*, *51*, 91-97.

Galfano, G., Rusconi, E., & Umilta, C. (2003). Automatic activation of multiplication facts: Evidence from the nodes adjacent to the product. *Quarterly Journal of Experimental Psychology*, *56A*, 31-61.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177-190.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*, 345-362.

Geary, D. C. (1996). The problem-size effect in mental addition: Developmental and cross-national trends. *Mathematical Cognition*, *2*, 63-93.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*, 4-15.

Geary, D. C., Bow-Thomas, C. C., Fan, L., & Siegler, R. S. (1993). Even before formal instruction, Chinese children outperform American children in mental addition. *Cognitive Development*, *8*, 517-529.

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development*, *67*, 2022-2044.

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology*, *27*, 787-797.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, *77*, 236-263.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*, 121-151.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*, 1343-1359.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, *74*, 213-239.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica*, *122*, 221-233.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*, B87-B95.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 32-44.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104-122.

Goldman, S. R., Mertz, D. L., & Pellegrino, J. W. (1989). Individual differences in extended practice functions and solution strategies for basic addition facts. *Journal of Educational Psychology*, *81*, 481-496.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology*, *93*, 615-626.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology, 82*, 375-386.

Hogan, T. P., & Brezinski, K. L. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning*, *5*, 259-280.

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*, 339-366.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*, 435-448.

Hung, Y.-H., Hung, D. L., Tzeng, O. J.-L., & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition, 106*, 1441-1450.

Imbo, I., & Vandierendonck, A. (2008a). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychological Research*, *72*, 331-346.

Imbo, I., & Vandierendonck, A. (2008b). Practice effects on strategy selection and strategy efficiency in simple mental arithmetic. *Psychological Research*, *72*, 528-541.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*, 662-673.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*, 103-119.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology*, *94*, 586-597.

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L. B., Felber, S., & Ischebeck, A. (2006). Neural correlates of the number-size interference task in children. *Neuroreport*, *17*, 587-591.

Keus, I. M., Jenks, K. M., & Schwarz, W. (2005). Psychophysiological evidence that the SNARC effect has its functional locus in a response selection stage. *Cognitive Brain Research*, *24*, 48-56.

Keus, I. M., & Schwarz, W. (2005). Searching for the functional locus of the SNARC effect: Evidence for a response-related origin. *Memory & Cognition*, *33*, 681-695.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1-23.

Koornstra, M. J., Neuwahl, N. M. E., & Van Hoorn, W. (1979). *IBO-Differentiatietest*. [Intelligence test for special education]. Lisse: Swets & Zeitlinger.

Lammertyn, J., Fias, W., & Lauwereyns, J. (2002). Semantic influences on feature-based attention due to overlap of neural circuits. *Cortex, 38,* 878-882.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*, 99-125.

Lee, K.-M., & Kang, S.-Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, *83*, B63-B68.

LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, *125*, 284-306.

LeFevre, J.-A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, *11*, 95-132.

LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 216-230.

Lemaire, P., Barrett, S. E., Fayol, M., & Abdi, H. (1994). Automatic activation of addition and multiplication facts in elementary school children. *Journal of Experimental Child Psychology*, *57*, 224-258.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children’s learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83-97.

Mabbott, D. J., & Bisanz, J. (2003). Developmental change and individual differences in children’s multiplication. *Child Development*, *74*, 1091-1107.

McCloskey, M., Harley, W., & Sokol, S. M. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 377-397.

McCloskey, M., & Lindemann, A. M. (1992). Mathnet: Preliminary results from a distributed model of arithmetic fact retrieval. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 365-409). Amsterdam: Elsevier.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, *74*, 240-260.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 46-60.

Montague, M., & Van Garderen, D. (2003). A cross-sectional study of mathematics achievement, estimation skills, and academic self-perception in students of varying ability. *Journal of Learning Disabilities*, *36*, 437-448.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.

Müller, D., & Schwarz, W. (2007). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Memory & Cognition*, *35*, 1151-1161.

Noël, M.-P., Rousselle, L., & Mussolin, C. (2005). Magnitude representation in children. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 43-54). New York: Psychology Press.

Opfer, J. E., & Thompson, C. A. (2006). Even early representations of numerical magnitude are spatially organized: Evidence for a directional magnitude bias in pre-reading preschoolers. *Proceedings of the 28th annual conference of the Cognitive Science Society*, 639-644.

Ostad, S. A. (1998). Developmental differences in solving simple arithmetic word problems and simple number-fact problems: A comparison of mathematically normal and mathematically disabled children. *Mathematical Cognition*, *4*, 1-19.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*, 348-367.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*, 55-78.

Pike, C. D., & Forrester, M. A. (1997). The influence of number-sense on children's ability to estimate measures. *Educational Psychology*, *17*, 483-500.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*, 416-442.

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, *21*, 387-399.

Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 250-257.

Rickard, T. C., & Bourne, L. E. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1281-1295.

Rickard, T. C., Healy, A. F., & Bourne, L. E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1139-1153.

Rosselli, M., Matute, E., Pinto, N., & Ardila, A. (2006). Memory abilities in children with subtypes of dyscalculia. *Developmental Neuropsychology*, *30*, 801-818.

Roussel, J.-L., Fayol, M., & Barouillet, P. (2002). Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology*, *14*, 61-104.

Rouselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*, 361-395.

Rubinsten, O., & Henik, A. (2005). Automatic activation of internal magnitudes: A study of developmental dyscalculia. *Neuropsychology*, *19*, 641-648.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology*, *81*, 74-92.

Ruijssenaars, A. J. J. M., Van Vliet, P. A. A., & Willemse, A. (2002). Het leren van rekenfeiten: Baart oefening kunst? Een verkennend onderzoek naar het automatiseren van de tafels van vermenigvuldiging [An explorational study of learning multiplication facts]. In A. J. J. M. Ruijssenaars & P. Ghesquière (Eds.), *Dyslexie en dyscalculie: Ernstige problemen in het leren lezen en rekenen* (pp. 165-181). Leuven, Belgium: Acco.

Rusconi, E., Galfano, G., Rebonata, E., & Umiltà, C. (2006). Bidirectional links in the network of multiplication facts. *Psychological Research*, *70*, 32-42.

Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction*, *1*, 217-244.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, *66*, 651-664.

Sekuler, R., & Mierkiewicz, D. (1977). Childrens judgments of numerical inequality. *Child Development*, *48*, 630-633.

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child & Adolescent Psychiatry*, *9*, II/58-II/64.

Shalev, R. S., & Gross-Tsur, V. (2001). Developmental dyscalculia. *Pediatric Neurology, 24*, 337-342.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*, 973-980.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250-264.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258-275.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428-444.

Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J.I.D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp 197-212). New York: Psychology Press.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy chioces in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*, 71-92.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237-243.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31-76). Hillsdale, NJ: Erlbaum.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*, 294-321.

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 7836-7841.

Treffers, A., Van den Heuvel-Panhuizen, M., & Buys, K. (Eds.) (1999). *Jonge kinderen leren rekenen. Tussendoelen annex leerlijnen. Hele getallen onderbouw basisschool.* Groningen: Wolters-Noordhoff.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*, 515-557.

Tzelgov, J., & Ganor-Stern, D. (2005). Automaticity in processing ordinal information. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 43-54). New York: Psychology Press.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 166-179.

Van den Bos, K. P., Zijlstra, B. J. H., & Spelberg, H. C. L. (2002). Life-span data on continuous-naming of numbers, letters, colors, and pictured objects, and word-reading speed. *Scientific Studies of Reading*, *6*, 25-49.

Van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology*, *87*, 239-266.

Van Harskamp, N. J., & Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication. Implications for the organisation of arithmetical facts. *Cortex*, *37*, 363-388.

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition*, *33*, 1-16.

Webster, R. E. (1979). Visual and aural short-term memory capacity deficits in mathematics disabled students. *Journal of Educational Research*, *72*, 277-283.

Wechsler, D., (2002). *The Wechsler Intelligence Scale for Children-III, Dutch version*. London: The Psychological Corporation.

Whalen, J., McCloskey, M., Lindemann, M., & Bouton, G. (2002). Representing arithmetic table facts in memory: Evidence from acquired impairments. *Cognitive Neuropsychology*, *19*, 505-522.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities*, *34*, 237-248.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*, 415-438.

Wood, G., Nuerk, H. C., & Willmes, K. (2006). Crossed hands and the SNARC effect: A failure to replicate Dehaene, Bossini and Giraux (1993). *Cortex*, *42*, 1069-1079.

Zbrodoff, N. J., & Logan, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 83-97.

Zebian (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, *5*, 165-190.

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., & Dong, Q. (2007). Chinese kindergartners' automatic processing on numerical magnitude in Stroop-like tasks. *Memory & Cognition*, *35*, 464-470.

Summary

This thesis focuses on the process of automatization in children’s learning of simple additions and multiplications. There are large individual differences in the ability to store arithmetic facts in long-term memory. Especially children with mathematical difficulties are often unable to solve simple arithmetic problems with retrieval from memory. The research presented in this thesis aims to answer several questions that are related to automatization in mathematics.

In the first study, automaticity in basic number processing was studied in 7-, 8-, and 9-year-old children, as well as in adults (Chapter 2). Two experimental tasks were used that are known to cause the so-called SNARC (spatial-numerical association of response codes) effect in adults. The SNARC effect refers to the finding that small numbers facilitate left responses, whereas larger numbers facilitate right responses. In the first task, number magnitude was essential to perform the task. In this task, a SNARC effect was found in all age groups. In the second task, number magnitude was irrelevant. In this task, a SNARC effect was found only in 9-year-olds and in adults. The findings are taken to suggest that (a) 7-year-olds represent number magnitudes in a way similar to that of adults and that (b) when perceiving Arabic numerals, children have developed automatic access to magnitude information by around 9 years of age.

It is often assumed that the ability to estimate magnitude and quantity is related to skill in mathematics. The development of numerical estimation and the relationship between numerical estimation and mathematical achievement was studied again in 7-, 8-, and 9-year-old children (Chapter 3). Four domains of numerical estimation were tested: number line, numerosity, length, and area. Principal components analysis revealed two different factors: (1) estimation of length and area, (2) number line and numerosity estimation. The two factors corresponded to different developmental trajectories over grades. Mathematics achievement scores correlated with 3 of the 4 estimation domains in 7-year-olds, but the relation between estimation and mathematics disappeared in higher age groups. It is concluded that estimation skill and general ability in mathematics are more intertwined in younger children than in older children.

In the three studies described next, the acquisition of arithmetic facts is studied with a learning task design. In Chapter 4, children practiced simple addition problems with three different methods: (a) writing down the answer, (b) choosing between two alternative answers, and (c) filling in the second missing addend. On a test with simple addition problems, children who practiced with the Choice method showed positive transfer: choosing between two answers was about as effective for learning addition facts as the conventional method of writing down the answer. There was no transfer effect for children who practiced with the Missing Addend method. The results are in accordance with network theories on

arithmetic fact learning and specifically the Identical Elements (IE) model of arithmetic fact representation. The IE model predicts no positive transfer when the numerical elements of a test problem do not match exactly with those of a practice problem.

The second practice study takes a closer look on transfer effects in children's arithmetic (Chapter 5). In two experiments, children practiced simple addition or multiplication problems. A positive transfer effect was found for problems with an operand order change; improvement was just as high as for practiced problems. No transfer effect was found for problems with one of the operands increased with one unit; improvement did not differ from problems unrelated to the practice problems. Analogous results were found for addition and multiplication, suggesting that storage and retrieval processes in both domains are highly similar in children.

The last practice study tries to find an explanation for the observation that some children seem to learn arithmetic facts almost automatically and other children keep struggling with basic addition and multiplication (Chapter 6). Recent research has shown that individual differences in working memory, counting speed, and rapid automatized naming are related to mathematical ability. In two experiments, children practiced simple addition or multiplication problems. The results showed that most children improved on the practiced problems. It was not possible to predict individual learning effects on the practice task from differences between children on measures of cognitive processing. However, correlational analyses revealed that Digit span forward, Digit span backward, Counting speed, and Rapid automatized naming seem to be related to mathematical ability. Furthermore, a domain-specific relationship was found between verbal short-term memory span and mathematical ability.

In the General discussion (Chapter 7), the findings of the research presented in this thesis are summarized and discussed.

Samenvatting

Dit proefschrift gaat over het automatiseringsproces in het leren van simpele optel- en tafelsommen, ook wel rekenfeiten genoemd. Er zijn grote verschillen tussen kinderen met betrekking tot het vermogen om rekenfeiten in het lange-termijn-geheugen op te slaan. Vooral kinderen met ernstige rekenproblemen zijn vaak niet in staat om simpele rekensommen op te lossen door het antwoord uit het geheugen op te halen. Het onderzoek in dit proefschrift probeert verschillende vragen te beantwoorden die te maken hebben met automatisering in het leren rekenen.

In het eerste onderzoek is automatisering in het verwerken van getallen onderzocht bij 7-, 8- en 9-jarigen en ook bij volwassenen (Hoofdstuk 2). Er werden twee taken gebruikt waarvan bekend is dat ze een zogenaamd SNARC-effect veroorzaken bij volwassenen. Het SNARC-effect is het gegeven dat op kleine getallen makkelijker met links gereageerd kan worden en op grotere getallen makkelijker met rechts. Om de eerste taak uit te kunnen voeren was het noodzakelijk om te weten hoe groot de getoonde getallen waren. In deze taak werd voor alle leeftijdsgroepen een SNARC-effect gevonden. Voor de tweede taak was getalgrootte niet relevant. In deze taak werd alleen een SNARC-effect gevonden bij 9-jarigen en bij volwassenen. De bevindingen duiden erop dat (a) 7-jarigen een zelfde soort representatie van getalgrootte hebben als volwassenen en (b) kinderen ongeveer vanaf hun negende automatisch toegang hebben tot getalgrootte na het zien van getallen.

Vaak wordt aangenomen dat er een relatie is tussen rekenvaardigheid en het schatten van grootte en hoeveelheden. De ontwikkeling van vaardigheid in schatten en de relatie tussen schatten en rekenen werd weer onderzocht bij 7-, 8- en 9-jarigen (Hoofdstuk 3). Er werden vier soorten schatopgaven getest: getallenlijnen, hoeveelheid, lengte en oppervlakte. Met een factor-analyse werden twee verschillende factoren gevonden: (1) schatten van lengte en oppervlakte, (2) schatten op de getallenlijn en schatten van hoeveelheid. Deze twee factoren waren ook terug te zien in de ontwikkeling van de scores in de verschillende leeftijdsgroepen. Bij de 7-jarigen waren de rekenscores gecorreleerd met 3 van de 4 typen schatopgaven, maar deze relatie tussen schatten en rekenen verdween in de hogere groepen. Geconcludeerd kan worden dat schatten en rekenvaardigheid meer aan elkaar gerelateerd zijn bij jonge kinderen dan bij oudere kinderen.

De drie volgende onderzoeken maakten gebruik van een leertaak-design om het verwerven van rekenfeiten te onderzoeken. In Hoofdstuk 4 oefenden kinderen simpele optelsommen met een van deze drie methoden: (a) het opschrijven van het antwoord, (b) het kiezen tussen twee mogelijke antwoorden, (c) het invullen van het tweede getal dat opgeteld moest worden. Na het oefenen kregen de kinderen een test met optelsommen waarvan ze het antwoord op moesten schrijven. Kinderen die geoefend hadden met de Keuze-methode lieten een generalisatie-effect zien op de test: het kiezen tussen twee antwoorden was ongeveer net zo effectief voor het leren van optelsommen als het oefenen met de reguliere methode waarbij het antwoord moest worden opgeschreven. De kinderen die geoefend hadden met de Invul-methode lieten geen generalisatie-effect zien. De resultaten zijn in overeenstemming met netwerktheorieën over het leren van rekenfeiten en in het bijzonder met het Identical Elements (IE) model voor de representatie van rekenfeiten. Het IE model voorspelt geen generalisatie-effect als de getallen van een som op een test niet precies overeen komen met de getallen van een som die geoefend is.

In het tweede leeronderzoek zijn generalisatie-effecten bij kinderen verder onderzocht (Hoofdstuk 5). In twee experimenten oefenden kinderen simpele optel- of tafelsommen. Er werd een generalisatie-effect gevonden bij sommen waarvan de getallen omgedraaid waren (6 + 3 geoefend, 3 + 6 in de test; 7 × 4 geoefend, 4 × 7 in de test): de kinderen gingen op deze sommen net zoveel vooruit als op de geoefende sommen. Als een van de twee getallen verhoogd was met 1 werd geen generalisatie-effect gevonden: er was geen verschil tussen de vooruitgang op deze sommen en vooruitgang op sommen die helemaal geen relatie hadden met de oefensommen. Voor optellen en vermenigvuldigen werden dezelfde resultaten gevonden, dus het lijkt erop dat bij kinderen in beide gevallen dezelfde soort opslag- en ophaalprocessen plaatsvinden in het geheugen.

In het laatste leeronderzoek is getracht om erachter te komen waarom sommige kinderen schijnbaar moeiteloos rekenfeiten leren en andere kinderen zoveel moeite blijven houden met simpele optel- en tafelsommen (Hoofdstuk 6). Uit recent onderzoek is gebleken dat individuele verschillen in werkgeheugen, telsnelheid en benoemsnelheid een relatie hebben met rekenvaardigheid. In twee experimenten oefenden kinderen simpele optel- of tafelsommen. Uit de resultaten bleek dat de meeste kinderen vooruitgingen op de geoefende sommen. Het was alleen niet mogelijk om individuele verschillen in de leeropbrengst op de oefentaak te voorspellen vanuit scores op cognitieve taken. Wel werd gevonden dat cijferreeksen nazeggen, cijferreeksen achterstevoren nazeggen, telsnelheid en benoemsnelheid gerelateerd waren aan rekenvaardigheid. Bovendien werd een domeinspecifieke relatie gevonden tussen de capaciteit van het verbale korte-termijn-geheugen en rekenvaardigheid.

In de Algemene discussie (Hoofdstuk 7) worden de bevindingen van het onderzoek in dit proefschrift samengevat en besproken.

Dankwoord

Tijdens mijn promotie-onderzoek heb ik steun gehad aan een heleboel mensen. De eerste persoon die ik wil bedanken is uiteraard Pieter Reitsma, mijn promotor en dagelijks begeleider. Pieter, bedankt voor je waardevolle adviezen, je geduld toen het allemaal wat langer duurde dan wij allebei hadden voorzien, en vooral voor het 'leren nadenken'. Verder wil ik de leescommissie bestaande uit Prof. dr. Marja van den Heuvel-Panhuizen, Prof. dr. Paul Leseman, Prof. dr. Ernest van Lieshout, Prof. dr. Han van der Maas en Prof. dr. Wied Ruijssenaars bedanken voor de aandacht die zij besteed hebben aan mijn proefschrift. Ernest van Lieshout wil ik daarnaast bedanken omdat ik af en toe bij hem langs kon komen om de opzet van een nieuw experiment door te spreken.

Dit onderzoek had niet uitgevoerd kunnen worden zonder de hulp van diverse afstudeerstudenten die een deel van de dataverzameling op zich hebben genomen. Janet Hospers, Baukje Hijlkema, Willeke Feenstra, Andrea van der Laan, Jolanda Geerts en Marloes Hunnekens: bedankt voor jullie inzet! Verder wil ik de 24 scholen bedanken die hebben meegewerkt aan het onderzoek. Dat zijn OBS Jules Verne, RK Basisschool Joannes XXIII, OBS De Kleine Dichter, OBS De Panda, OBS De Cirkel, Daltonschool Jan Nieuwenhuyzen, Jenaplanschool Cleophas en PC Basisschool Op Dreef uit Utrecht. OBS De Dapper, St. Lidwinaschool en OBS De Kaap uit Amsterdam. RK Basisschool De Cirkel en OBS Martin Luther King uit Amstelveen. OBS De Syncope, PC Basisschool De Lichtboei en Oec. Basisschool De Omnibus uit Almere. PC Basisschool De Sprong en PC Basisschool De Weidebloem uit Hoogeveen. PC Basisschool De Regenboog uit Noordscheschut. OBS Van Kleffens en OBS De Commanderije uit Heerenveen. OBS Beatrixschool uit Steenwijk. St. Antoniusschool uit Kortenhoef. En tenslotte CBS Wakersduin uit Noordwijk.

Binnen PI Research waren er diverse mensen waar ik altijd een beroep op kon doen. Jan Geelhoed was altijd geïnteresseerd in hoe mijn onderzoek verliep en nam vaak de moeite om even met me mee te denken. Hedy Stegge heeft me geholpen om mijn ingewikkelde regressie-analyses uit te pluizen. Hans van Muijen was een trouwe bondgenoot in het bedwingen van problemen met mijn computer en het netwerk. En bij Louise Stavenuiter kon ik terecht voor alle dagelijkse vragen en problemen. Verder wil ik Alexandra Benschop van de Pionier en Adriaan Kievit van de Bascule bedanken voor het uitlenen van hun materialen voor mijn experimenten. Grote dank ben ik verschuldigd aan Paul Groot van de VU die zich niks aantrok van het feit dat de VU-bureaucratie mij niet als officiële werknemer zag en altijd binnen no time mijn E-Prime problemen wist op te lossen.

Mijn tijd op het PI was lang niet zo leuk geweest zonder mijn gezellige collega's. Na de leegloop in de academische kernen zou je het niet zeggen, maar ik heb in de loop van de jaren behoorlijk wat directe collega's gehad. Afke, Annelise, Eduardo, Else, Inez, Joke, Judith, Maartje, Mieke, Miranda, Siety, Tessa en Vincent bij mijn eigen Leerstoornissen-afdeling en Evelien, Marianne, Marieke, Marjolein, Patty, Sander Begeer, Sander Thomaes en Yoast bij de overige academische kernen. Lieve collega's: ik denk met veel plezier terug aan onze net iets te lange lunches buiten in de zon, de discussies die liepen van de laatste film tot de wetenschapsquiz, de chocolademelk- en koekjesmomenten en de eetafspraken die nog geregeld voorkwamen nadat er al veel mensen van het PI waren vertrokken. Daarnaast wil ik Afke, Annelise en Inez bedanken voor al hun adviezen bij het opstarten van mijn onderzoek. In het bijzonder wil ik Maartje bedanken, die ondanks de lonkende lege kamers altijd mijn trouwe kamergenoot is gebleven. Maartje, je onuitputtelijke optimisme was erg inspirerend voor mij. Bovendien was je de ideale kamergenoot, want we konden stil en geconcentreerd werken als dat nodig was, maar als ik behoefte had aan een praatje was daar ook altijd ruimte voor. Zowel onze inhoudelijke gesprekken over de dingen waar ik tegenaan liep in mijn analyses als de gezellige cola-momenten in het dooie punt van de middag waren volgens mij cruciaal voor de voortgang van mijn onderzoek! Met Gonnie, Marjan, Tessa en Marjolein heb ik in de OR gezeten van PI Research, wat ik leerzaam maar ook heel gezellig vond. En met Yoast en Marjolein heb ik een onvergetelijke PIR Doedag georganiseerd met zon, zee, strand, vliegers en sushi.

Ik had niet alleen directe collega's bij PI Research, maar ook collega's verder weg. Eva, Judith en Femke van de UvA maakten elke onderlinge bijeenkomst gezellig en dankzij hen weet ik nu bijna wat het verschil is tussen fonologie en morfologie... Met Vanessa van de UvA heb ik de cursus Multilevel-analyse tot een goed einde gebracht (en dat is behoorlijk ingewikkeld!). Evelien Dirks van de VU hield me op de hoogte van haar onderzoek en gaf me het gevoel dat ik niet de enige AiO was die iets met rekenen deed. Titia van de UU heeft me meegesleept naar een bijeenkomst over getalverwerking in Aken (in het bizarste gebouw dat ik ooit heb gezien!) die een enorme inspiratie vormde voor mijn SNARC-artikel. Frank van de UU werd van een vriend ineens ook een collega en hielp me op weg met diverse statistische analyses. En tenslotte heb ik het gezelschap van Fenna, Sylvia en Iris van het FI erg op prijs gesteld: tijdens onze afspraken dwaalden de gesprekken al snel af van het AiO-lief-en-leed naar onze laatste vakanties en dat is natuurlijk helemaal niet erg!

Heel belangrijk voor het slagen van een promotie-onderzoek is niet alleen het harde werken, maar ook de ontspanning. Mijn lieve vriendinnen en paranimfen Judith en Eva hebben uiteraard al mijn klachten geduldig aangehoord, maar er ook voor gezorgd dat ik eens wat leuks ging doen als ik het druk had. Zoals de stad ingaan om kleren te kopen, of was ik niet degene die dat leuk vond? Een bijzondere plaats in de afgelopen jaren was er voor Ela, dochter van Judith en Jeroen, die ik met veel plezier heb zien opgroeien. Ela is bijna net zo oud als mijn proefschrift, maar wat kan een kind veel leren in 4,5 jaar! Ontspanning was er ook als ik bij Daniel koffie ging drinken en zag wat Marijn nu alweer allemaal kon, als ik met Paul naar de sneak preview ging en op mijn donderdagavond-

repetities bij Wijkorkest de Noordooster. En laat ik Geert van de fitness niet vergeten die af en toe mijn nek uit de knoop haalde als ik te gespannen achter de computer had gezeten.

Mijn familie heeft mijn inspanningen van een afstandje denk ik met wat verbazing gade geslagen. Jop, Tesja en Merel: ik weet niet of ik er echt in geslaagd ben om uit te leggen waar ik mee bezig was, maar je kunt natuurlijk altijd nog een poging wagen om mijn proefschrift te lezen… Mijn ouders, Pauline en Frans, wil ik bedanken voor hun inhoudelijke adviezen (wel apart natuurlijk, een dochter die promotie-onderzoek doet op het snijvlak van jullie eigen werk!) en omdat ze ongetwijfeld meer vertrouwen in mij hadden dan ikzelf. Mijn schoonfamilie, Henny, Lita, Barry en Rima, werd het nooit moe om te vragen hoe het met mijn onderzoek ging. Maar mijn grootste dank gaat natuurlijk uit naar mijn geliefde. Lieve Martin, ik maakte wel eens een grapje dat ik al niet genoeg bladzijden in mijn proefschrift zou hebben om je te bedanken voor je technische ondersteuning. En dat is natuurlijk zo, je hebt me altijd geholpen met alle computerproblemen en was onmisbaar in de uitvoering van mijn experimenten. Maar ik wil je toch vooral bedanken voor je steun, liefde en vertrouwen. Zonder jou had ik het niet gekund.